Fantastic Pa and Where to Find Them: A Spatial Analysis of the pre-European Pā of Aotearoa

by
Baylee Ann Smith

A thesis submitted in fulfillment of the degree of Master of Arts

in the Department of Anthropology and Archaeology

December 2017
Abstract

This thesis presents a spatial analysis of pre-European Pā sites across New Zealand with a focus on the North Island. Few spatial analyses of Pā have utilised sophisticated statistical methods or been carried out at a large scale. Those that have largely do not account for potential confounders or investigate the implications of their results in terms of social organisation. These past studies therefore motivate our three research questions: What is the density distribution and clustering pattern of Pā sites in New Zealand? What are the dominant variables governing the density and clustering of Pā sites across New Zealand? What are the implications of the identified patterns in terms of social organisation?

Kernel density estimates are calculated for the Pā of the North and South Island. These estimates support previous statements by researchers about the distribution of Pā: Pā are most dense in the north of the North Island, decreasing southwards. The density of Pā is then compared to the density of non-Pā via the relative risk function in order to curb the confounding effects of population and bias introduced by unsystematic surveying. This shows that Pā density is relatively high through the central North Island and relatively low through Auckland and Northland.

Clustering is explored using the pair-correlation function (PCF). This reveals that Pā show strong evidence of large clusters in the absence of deterministic heterogeneity. Heterogeneity in the Pā point pattern from sources such as geography and the human population distribution is then accounted for by utilising the inhomogeneous PCF. This suggest that Pā in the North Island form clusters that then repel each other. The central and northern regions identified by the relative risk are compared: the central North Island shows the same pattern as the North Island in general but the northern region demonstrates a second level of clustering at larger distances indicating “clusters of clusters.”

The spatial relative risk of Pā is modelled against temperature, solar radiation, land wetness, distance from the coast, soil drainage class, and slope. In all the models fitted only the distance to coast parameter is significant. The parameter estimate indicates that the relative risk of observing a Pā increases as the distance from the coast increases.

The spatial patterns identified are interpreted as indicating that society and by extension Pā were centred around the hapū, which were largely independent of their respective iwi, except in the northern North Island where there was more political centralisation. This resulted in a proportionately small number of large Pā in the north and large number of small Pā through the central North Island. More neighbours through the central region may have also promoted more Pā building.

We conclude that Pā are driven more by social factors than environmental ones.
Acknowledgements

To my incredible family for their unwavering love and support, especially in difficult times.

To my fellow anthropology and archaeology postgraduate students for sharing this journey with me.

To my pole family for consistently motivating me to actually leave my office in the evenings and get some physical activity in. Without you I likely would have become a hermit.

To Alice, Harry and Morgan, even if we struggle to find time to physically spend together.

To Aaron, for being kind, patient, selfless, and supportive, for tempering my self doubt, and for being my best friend.

To Kevin Jones for supplying the aerial photographs used in this text and for being so forthcoming with help and advice.

Finally, to my supervisors, Dr Tim Thomas and Dr Tilman Davies for their expertise and guidance. Without you this would not have been possible.

Thank you.
Abstract i

Acknowledgements ii

List of Figures v

List of Tables vii

Abbreviations viii

1 Introduction 1

1.1 Thesis Outline 3

2 Approaches to Pā studies 6

2.1 What is a Pā? 6

2.1.1 The definition of Pā 7

2.1.2 Classification of Pā 8

2.1.3 The Trigger for Pā Construction 9

2.2 The Function of Pā 10

2.3 Pā Distribution 14

2.3.1 Social Organisation 17

2.4 Conclusions 18

3 The Data 20

3.1 A Brief History of Site Recording in NZ 20

3.2 Implications for the Following Analyses 23

3.3 Our Datasets - Filtering the Data 24

4 Exploratory Analysis: Pā Density 29

4.1 Background: Spatial Statistics 29

4.1.1 Point Pattern 29

4.1.2 Point Process 30

4.1.3 Intensity and Density 30
4.2 Density Estimation: Method ........................................ 32
4.3 Density Estimation: Results ...................................... 35
4.4 Density Ratio: Method .............................................. 38
4.5 Density Ratio: Results .............................................. 43
4.6 Conclusions .......................................................... 45

5 Exploratory Analysis: Pā Clustering .......................... 51
5.1 Background: Spatial Statistics ................................. 52
  5.1.1 Complete Spatial Randomness, Homogeneity and
         Heterogeneity .................................................. 52
5.2 Pair-Correlation Function: Method ......................... 53
5.3 Pair-Correlation Function: Results ......................... 55
5.4 Inhomogeneous Pair Correlation Function:
         Method .......................................................... 58
5.5 Inhomogeneous Pair Correlation Function:
         Results .......................................................... 59
5.6 Conclusions .......................................................... 62

6 Modelling the Distribution of Pā ............................... 64
6.1 Background: Statistical Modelling ......................... 64
  6.1.1 Generalised Least Squares ................................. 66
6.2 Geostatistical Modelling: Method ......................... 67
6.3 Covariate Selection .............................................. 72
  6.3.1 Water Bodies .................................................. 73
  6.3.2 Ridges .......................................................... 74
  6.3.3 Agriculture .................................................... 76
6.4 Results - Single Covariates .................................. 81
6.5 Results - Multiple Covariates .............................. 84
6.6 Conclusions .......................................................... 86

7 Discussion and Conclusions ..................................... 88
7.1 Looking to the Future ............................................. 94

References Cited ....................................................... 96
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Aerial photograph of Pouerua Cone Pā</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Aerial photograph of Taniwha Pā</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>The three “agricultural zones” of New Zealand</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>NZAA Site Record Form</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Locations of pre-European Pā in the North Island</td>
<td>26</td>
</tr>
<tr>
<td>3.3</td>
<td>Locations of pre-European Pā in the South Island</td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>Locations of pre-European non-Pā</td>
<td>28</td>
</tr>
<tr>
<td>4.1</td>
<td>Map of excavated Phase Four Bronze Age burials at Ban Non Wat</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>Illustration of the kernel density estimate in action for 1-dimensional data using four different bandwidths</td>
<td>33</td>
</tr>
<tr>
<td>4.3</td>
<td>Illustration of the kernel density estimate in action for 2-dimensional point pattern data</td>
<td>34</td>
</tr>
<tr>
<td>4.4</td>
<td>Kernel-smoothed estimate of the probability density function of Pā sites in the North Island of New Zealand</td>
<td>36</td>
</tr>
<tr>
<td>4.5</td>
<td>Kernel-smoothed estimate of the probability density function of Pā sites in the South Island of New Zealand</td>
<td>37</td>
</tr>
<tr>
<td>4.6</td>
<td>Kernel-smoothed estimate of the probability density function of non-Pā sites in the North Island</td>
<td>40</td>
</tr>
<tr>
<td>4.7</td>
<td>Kernel-smoothed estimate of the probability density function of non-Pā sites in the South Island</td>
<td>41</td>
</tr>
<tr>
<td>4.8</td>
<td>Fixed-bandwidth relative risk surface for Pā sites compared with non-Pā sites in the North Island</td>
<td>44</td>
</tr>
<tr>
<td>4.9</td>
<td>Fixed-bandwidth relative risk surface for Pā sites compared with non-Pā sites in the South Island</td>
<td>45</td>
</tr>
<tr>
<td>4.10</td>
<td>Adaptive-bandwidth relative risk surface for Pā sites compared with non-Pā sites in the North Island</td>
<td>46</td>
</tr>
<tr>
<td>4.11</td>
<td>Adaptive-bandwidth relative risk surface for Pā sites compared with non-Pā sites in the South Island</td>
<td>47</td>
</tr>
<tr>
<td>4.12</td>
<td>Fixed-bandwidth relative risk surfaces for Pā sites compared with six non-Pā site types in the North Island</td>
<td>49</td>
</tr>
<tr>
<td>5.1</td>
<td>Two hypothetical examples of 2D point patterns</td>
<td>53</td>
</tr>
<tr>
<td>5.2</td>
<td>Pair-correlation function for Pā sites in the North Island</td>
<td>56</td>
</tr>
<tr>
<td>5.3</td>
<td>Pair-correlation functions for the northern and central relative risk regions</td>
<td>57</td>
</tr>
</tbody>
</table>
List of Figures

5.4  Inhomogeneous pair-correlation function for Pā in the North Island .......................... 59
5.5  Inhomogeneous pair-correlation function for Pā in the North Island .......................... 60
5.6  Inhomogeneous pair-correlation function for Pā in the North Island .......................... 61
5.7  Inhomogeneous pair-correlation functions for the northern and central relative risk regions ............................................................... 62

6.1  Results of Leathwick (2000) ................................................................. 65
6.2  Two point patterns randomly generated from an inhomogeneous Poisson point process ................................................................. 68
6.3  Estimated relative risk for the two randomly generated point patterns ................................................................. 69
6.4  Variogram plot showing the estimated semivariance of the example relative risk with line of best fit ................................................................. 71
6.5  Percentage of pixel area covered by freshwater bodies ................................................................. 75
6.6  Average slope in degrees ................................................................. 76
6.7  Mean annual solar radiation in megajoules per square metre per day, multiplied by 10 ................................................................. 78
6.8  Mean annual temperature in degrees Celsius, multiplied by 10 ................................................................. 79
6.9  Soil drainage class ................................................................. 80
6.10 Parameter estimates with corresponding 95% confidence intervals at a range of bandwidths for parameters water, sol, slope, temp and dist ................................................................. 82
6.11 Parameter estimates with corresponding 95% confidence intervals at a range of bandwidths for each level of the soil parameter ................................................................. 83
7.1  Aerial photograph of One Tree Hill/Maungakiekie Pā ................................................................. 92
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Correlation parameters for our example dataset estimated using variogram modelling.</td>
<td>70</td>
</tr>
<tr>
<td>6.2</td>
<td>Parameter estimates for our example dataset following generalised least squares.</td>
<td>72</td>
</tr>
<tr>
<td>6.3</td>
<td>Summary of covariates to be used in geostatistical modelling of relative risk.</td>
<td>81</td>
</tr>
<tr>
<td>6.4</td>
<td>Parameter estimates and corresponding P-values for the geostatistical model containing all six parameters.</td>
<td>84</td>
</tr>
<tr>
<td>6.5</td>
<td>Interaction estimates and P-values.</td>
<td>85</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>2 Dimensional</td>
</tr>
<tr>
<td>CSR</td>
<td>Complete Spatial Randomness</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographical Information System(s)</td>
</tr>
<tr>
<td>KDE</td>
<td>Kernel Density Estimate</td>
</tr>
<tr>
<td>NZAA</td>
<td>New Zealand Archaeological Association</td>
</tr>
<tr>
<td>OS</td>
<td>Over Smoothing bandwidth</td>
</tr>
<tr>
<td>PCF</td>
<td>Pair Correlation Function</td>
</tr>
<tr>
<td>RR</td>
<td>Relative Risk</td>
</tr>
</tbody>
</table>
For Justin
Chapter 1

Introduction

Spatial analysis is an increasingly popular and complex field of archaeology. Its origins date back to some of the earliest modern archaeological research around the world. In British archaeology, an interest in spatial patterns is seen as far back as Pitt Rivers in the late 19th century (Daniels 1950), as well as in the work of, for example, Fox (1923) and Ward-Perkins (1955). In the Americas the first glimpse of a spatial approach to archaeology is seen in the work of Morgan in the 1880s (Philips & Campbell 2004). The spatial arrangements of artefacts, features, and monuments were often a focus of these early studies, which explored settlement patterns and environmental explanations for observed spatial patterns. These early studies were also largely descriptive, often focusing on building culture history narratives rather than attempting to explain observed pattern in terms of social organisation.

In the last thirty to forty years, however, we have seen a paradigm shift in approaches to spatial archaeology; this research is now more concerned with how people’s use of space, as well as the spatial patterns this creates, reflects ecological, economic, social, political, ideological, and other facets of human culture. This shift was influenced in large part by the work of Gordon Willey (1953), who is regarded as pioneering the settlement pattern approach. His work in the Viru Valley of Peru included analysis of the distribution of archaeological sites and how this reflected human behaviours and how they interacted with their environment.

While more interpretive, these spatial analyses were often still very basic with regards to their methodologies, consisting largely of the inspection of artefact or site location maps, with inferences drawn about past human behaviours based on perceived spatial patterns. The use of the word perceived here is important; such analyses were often highly subjective, and therefore could only offer limited (or in
the worst cases, misleading) insight into past behaviours.

In the 1960s and 1970s processualists started to adopt and adapt models from geography and ecology to explain archaeological spatial patterns. This culminated in the two seminal spatial archaeology texts Hodder & Orton (1976) and Clarke (1977), which set the standard for more formal analyses of archaeological spatial data grounded in statistical testing that curbed the subjectivity of earlier approaches. This shift to a scientific approach to answering spatial questions also allowed researchers to cope with the increasing number of sites being recorded.

Technological improvements in recent years have advanced the field of spatial archaeology even further by allowing the development of powerful geographical information systems (GIS) and making more sophisticated statistical analyses possible. Employing these methods, common themes of spatial archaeology research have emerged including ecological relationships, social patterning, political organisation, landscape construction and perception, and monumentality.

The development of a spatial approach to archaeological questions has followed a similar trajectory in New Zealand. Spatial archaeology has long been a central field of archaeological research in New Zealand with settlement pattern analysis being one of its most common and enduring themes. Some of New Zealand’s earliest archaeological studies have considered spatial patterns of sites (e.g. Best 1918a,b, Skinner 1921, Adkin 1948). Spatial analysis became a more formalised approach to archaeology when Willey’s student Roger Green introduced his settlement pattern analysis to New Zealand. This began with his influential study of the Auckland province (Green 1963, 1970) where he proposed a prehistoric cultural sequence with each phase marked by a distinctive settlement pattern. This study formally introduced settlement pattern analysis to New Zealand and laid the methodological framework for many future settlement pattern studies, particularly the combination of different types of evidence (environment, artefacts, subsistence, settlement, etc.). Other influential works include Groube (1964, 1965), and Triger (1968), all of which explored spatial patterns of archaeological sites. As with spatial archaeology in the rest of the world, analyses of the 1960s and 1970s introduced more formal and scientific methods for analysing spatial patterns; in many ways, however, spatial archaeology in New Zealand has not evolved far beyond this. Although resources like GIS are now widely utilised, the sophisticated spatial statistics methods now available are rarely applied to spatial data in New Zealand archaeology.
Today, New Zealand has a wealth of recorded archaeological sites at sample sizes ideally suited to carrying out spatial analyses, particularly at a large scale, and the field of spatial statistics provides a great tool kit for analysing such datasets. Pā sites in particular are well suited to large scale spatial analyses. At present there are a little over 7,000 Pā recorded across New Zealand. Their large size and enduring place in local Māori traditions means that these recorded sites likely represent a large proportion of the Pā that have been built by the Māori of Aotearoa and therefore their spatial arrangement should be informative. For these reasons, this site type and the analysis of its distribution across New Zealand using the latest spatial statistics methods forms the main focus of this thesis.

When a spatial distribution is non-random, as is the case with Pā sites, we must assume that there is some factor or factors driving this non-random pattern. Often in archaeology, researchers attempt to explain such patterns by searching for correlations. However, as the adage goes, correlation does not necessarily imply causation. Despite this, spatial correlations between archaeological sites and certain factors are often uncritically interpreted as evidence of a causal relationship where this factor or factors dictates the spatial pattern of the sites of interest. With regards to settlement pattern studies in New Zealand a reliance on topographic data has resulted in many environmentally deterministic models being adopted based on correlations between sites and environmental factors. In many cases correlations may in fact be driven by confounders. Confounders are variables that influence two or more other variables, creating spurious associations between them; they therefore need to be identified and accounted for in order to identify true causal relationships. The issue of confounders has not been adequately addressed in spatial archaeology in New Zealand. This thesis will therefore employ methods drawn from spatial statistics in order to explore the distribution of Pā and factors that contribute to it, that account for possible confounders.

1.1 Thesis Outline

This thesis will proceed as follows:

Chapter 1 is the current chapter and provides a brief introduction to the topic of this thesis, and describes the theoretical background that motivates our study. This chapter also provides an overview of the content of each chapter.

Chapter 2 presents a summary and critique of the diverse range of archaeological research that has been conducted on Pā in New Zealand to date. This
Chapter 1 Introduction

literature review is organised into three sections, each focusing on a major theme of Pā research. The first section focuses on the debate regarding what constitutes a Pā, looking at proposed definitions for Pā, classification systems, and theories around the trigger(s) for Pā construction. Section 2 explores the varied opinions on and evidence for the function(s) of Pā with four main schools of thought analysed: Pā as symbolic structures, Pā as defended settlements, Pā as citadels, and Pā as defended food stores. The third and largest section explores the main focus of this thesis: Pā distribution. This section summarises the findings of previous spatial analyses at regional and national scales and what they tell us about pre-European Māori social organisation. The key lessons from this literature review are then summarised and used to finalise and motivate our research questions, the focus of the remainder of this thesis.

Chapter 3 describes the main dataset that will be employed in the analyses described in this thesis. It begins with a brief overview of site recording in New Zealand and how this impacts the quality of our data. We then describe the filtering process that was used to “clean” our data to make it appropriate for subsequent statistical analyses.

Chapter 4 explores the density of Pā across New Zealand. It opens with a critique of methods often used by archaeologists to describe the distribution or density of Pā then introduces some important statistical concepts (point pattern, point process, intensity and density) before describing our own methods: the kernel density estimation procedure, and density ratios. The results of these two methods are described and their implications discussed.

Chapter 5 looks at evidence for clustering of Pā and follows the same structure as chapter 4. Previous methods used to explore clustering are critiqued, important statistical concepts are introduced (complete spatial randomness, heterogeneity, homogeneity), and our own methods, the pair correlation function and inhomogeneous pair correlation function, are described. Finally, the results of this analysis are described and the implications thereof are discussed.

Chapter 6 focuses on modelling the spatial distribution of Pā. This chapter also opens with a brief critique of approaches that researchers have previously used to model the Pā distribution as well as an overview of generalised least squares model fitting. We then provide a simple example to illustrate the geostatistical methods that will be employed. This is followed by a discussion of how each of our model covariates were selected accompanied by a description of the datasets used for each covariate. The results of our models are then given, divided into two
sections: one for single covariate models, and one for multiple covariate models. The implications of these results are then discussed.

The seventh and final chapter provides a summary of the results of all of the analyses presented in this thesis followed by an in-depth discussion of the implications of these results. Possible explanations for patterns observed are debated and hypotheses about pre-European Māori social organisation are presented based on these patterns. Finally we offer some suggestions about the direction of future research into the distribution of Pā in New Zealand.
Chapter 2

Approaches to Pā studies

This chapter consists of a formal review of the wide expanse of literature on the Māori Pā in order to motivate and formulate our research questions. The first section will synthesise literature on the definition of Pā and attempts to categorise these sites as well as potential triggers of Pā-building; the second section will look at the function of the Pā and its role in warfare; the final section will focus on the core issue of this thesis: the distribution of Pā and their role in the study of pre-European settlement patterns and social organisation.

2.1 What is a Pā?

The term ‘Pā’ is a Māori word meaning a ‘fortified place’ or ‘enclosure’ (Best 1927, 18) and is used by archaeologists to refer to a group of sites that have defensive features: typically, but not always, earthworks. However, there is no standard definition of what defensive features are necessary for a site to be considered a Pā. This is illustrated clearly in ArchSite (the national database of archaeological sites in New Zealand) where the sites that have been recorded as Pā vary significantly with regards to the nature of the defensive features that they consist of. These features can range from natural defences such as cliffs or single terraces to complicated ditch and bank arrangements, or may be a combination of these. This is because ArchSite is not a site classification system and lacks any standard criteria for distinguishing a Pā site; these sites therefore can vary significantly morphologically while still being classed as the same site type. This is an issue for archaeologists as the results of studies involving sites classified as Pā may be inconsistent if, as Schmidt (1996a) suggests, the designation of any given site as a Pā is going to be based on the interpretation of the first person to record it.
Compounding this issue is the fact that very few Pā studies offer a formal definition of what the author considers to be a Pā or what their selection criteria for obtaining sites for analysis was before proceeding. Instead, there appears to be an assumption that the reader already understands what a Pā is, which may not be reasonable. It is concerning that while Pā have been a central focus of New Zealand archaeology for so long we still lack the basic foundation of what a Pā is for these studies. No analysis as of yet has addressed how these inconsistencies may be affecting Pā studies. Such a study may be very valuable, but is outside the scope of this thesis.

Archaeologists have previously tried to solve this issue through devising a formal archaeological definition of Pā, and through classification systems.

### 2.1.1 The definition of Pā

Although Mihaljevic (1973, 139) argues that producing an acceptable, archaeological definition for Pā is “One of the greatest problems of New Zealand archaeology”, this issue has received proportionately little attention in the literature with few archaeologists offering candidates for a formal definition. Two early attempts at a working definition for Pā were made by Alistair Buist (1964, 20) and Kenneth Gorbey (1970, 27) in order to identify Pā sites for their own spatial analyses. Buist (1964, 20) used the strict definition of “an area of land enclosed by a ditch or a ditch and bank or a scarp” for his study of Northern Taranaki Pā. This definition may have only limited usefulness as it does not allow for combinations of ditches, banks, and scarps, or for Pā that make use of natural defences, which are fairly common in Taranaki (Prickett 1982, 1983). Gorbey in his spatial analysis of Pā from the whole of New Zealand, offered a more accommodating three part definition to describe Pā:

1. An area of land enclosed by
   
   (a) a bank and ditch system
   
   (b) a scarp system

2. Terrace arrangements rise to a seemingly easily defended area

3. The site was an artificial swamp mound.
Buist’s definition has been employed uncritically by Wilkes (1995, 239) for recording Pā sites on the King Country coastline, but aside from this these two definitions have not been utilised any further.

The closest to a formal definition that has been used comes from the New Zealand Archaeological Association handbooks for recording sites. The 1979 handbook offers the somewhat vague definition: “A site containing built fortifications” (Daniels et al. 1979, 24). This definition could be criticised for not distinguishing between palisades and earthwork fortifications; under this definition a palisaded village could be considered a Pā. The 1999 site recording handbook (Walton 1999, 47–53) notes that “Pa were often built on hills and ridges ... The common identifying features of pa were earthwork defences (ditches and banks ...) and frequently palisading ... Palisades do not usually survive in an archaeological context so the main field identifier is now earthworks.” The definitions from these guidebooks arguably had little effect on what sites were being recorded as Pā. For example, neither of these definitions allows for natural defences such as cliffs yet many naturally defended sites have been recorded as Pā in ArchSite. These definitions also do not mentions scarps or terraces.

Despite these options, no formal definition has as of yet been adopted and applied consistently by archaeologists. Anderson (2009) argues that this has led to the distinction between Pā and lightly palisaded or undefended settlements becoming blurred. Because archaeologists are not using a single definition for Pā it could be that different studies are using different subsets of Pā that may ignore or include sites inappropriately. Examining these varied definitions also reinforces that archaeologists should not assume that the reader knows what a Pā is (or rather what the researcher considers to be a Pā) when publishing their research.

### 2.1.2 Classification of Pā

In the absence of a useful objective definition for Pā, classification systems have been suggested as a means of coping with the variation that had made producing a formal definition of Pā difficult (Mihaljevic 1973, 143).

One of the earliest classification systems was put forward by Jack Golson (1957) to assist archaeologists in identifying and recording sites for the NZAA site recording scheme. This classification was largely morphological, distinguishing lowland Pā from flat land Pā and then further dividing these into subclasses based on the types of earthwork defences present (terraces, transverse ditches and banks, and transverse and lateral ditches and banks). In contrast, the classification system
put forward by Buist (1965), based on his surveys in the Taranaki region, focused more upon the social complexity of Pā sites by using the number and types of 'units' - distinct living platforms - as an analogue.

The most popular classification system is that of Les Groube (1970). This system bases its three classes on the common morphological features of Pā: terraces, transverse ditches, and parallel ditches. A fourth class was added to this system by Aileen Fox (1976, 16) to include swamp and lowland Pā. Davidson (1987, 16–7) warns that this classification system should be used cautiously as it was designed for use in culture history studies, but may be useful for descriptive purposes (e.g. Schmidt 1996a).

The NZAA site recording handbooks (Daniels et al. 1979, Walton 1999) do advise that such classification systems may be useful for identifying and recording Pā sites, however, these systems dropped off in popularity due to the difficulty in applying categories meaningfully. Despite their drawbacks, these classification systems do highlight the important issue of variation in Pā form and bring into question whether it is in fact appropriate to group all types of Pā together under a single site type.

### 2.1.3 The Trigger for Pā Construction

By analysing radiocarbon dates from Pā sites across the country, Schmidt (1996b) determined that the commencement of Pā building occurred by 1550 CE, beginning around 1500 CE, with no one area seeing the initiation of Pā building. There has been little consensus between archaeologists on what they believe to be the trigger for this seemingly sudden emergence of Pā building in Aotearoa; no other oceanic island chain saw fortifications spread in such a short time or in such numbers (Barber 1996). Understanding what caused this sudden proliferation may be important in the interpretation of Pā and their distribution.

Groube (1970, 2) has argued that because Pā construction was so labour intensive, requiring high per capita involvement, they must not have been built until they were needed. The most commonly posited theory is that Pā emerged as the result of a growing population size and the subsequent competition for good soil, particularly for kumara cultivation (Allen 1994, Buist 1964, Davidson 1984, 1987, Duff 1967, Gorbey 1970, Sullivan 1985, Vayda 1956, 1960, 1961). It has been argued that this is why we see so few Pā in the South Island; the relative weakness of the economic base and small population size did not necessitate their construction (Walton 2001). Graves & Sweeney (1996) and Hunt & Lipo (2001)
similarly proposed that population increase was the trigger for Pā building but, in contrast to these earlier archaeologists, argued that Pā construction was used as a means to divert energy from agricultural production and thus suppress the rising population and the competition that accompanied it. Conversely, Groube (1970) argues that any man/land ratio-centred argument is difficult to maintain within the narrow temporal and extensive geographical boundaries of pre-European New Zealand, which are further constrained by the limited ability of the subsistence economy of the time to support Pā construction. Offering further support to this hypothesis, some have argued that there was enough cleared arable land to sustain the maximum pre-European population and thus preclude competition for this resource (Davidson 1984, 182). In addition, Gorbey (1970) points out that there were extensive tracts of cultivable land in areas such as the Waikato that were not cleared and settled. He thus questions how population pressure was great enough to form the impetus for the construction of fortifications, but not great enough to compel people to tackle such land.

An alternative hypothesis that has been suggested by Buck (1949) and Mead (1975) is that the Pā is a manifestation of tribal identity. That is, as the importance of tribal identity grew, so too did the need to protect the tribal lands, hence creating the impetus for the building of fortifications. Barber (1996, 877) likewise highlights the importance of tribal identity, arguing that Pā “may be correlated with the socio-political development of the ancestral landscape and, eventually, the formation of new founding traditions and descent associations.” Davidson (1987) has argued that such hypotheses are not amenable to archaeological testing. However, we argue that such behaviours may leave spatial markers that could be analysed.

2.2 The Function of Pā

The often conspicuous earthworks that make up Pā have been interpreted by many archaeologists as evidence of warfare, particularly in the so-called “Classic Period” (Allen 2006, Duff 1967, Green 1967, Walton 2001). However, Vayda (1960, 83) has argued that we must be cautious in making such conclusions; Pā building is not the same thing as warfare and there is currently no clear indication about the frequency of pre-European fighting as so few Pā sites (or battle sites in general) have been excavated Schmidt (1996b). This means we have little direct evidence of warfare (or lack thereof) at Pā or other sites.
An alternative interpretation of Pā seen in the literature is as a symbolic structure. Mihaljevic (1973) discounted the military aspects of Pā in favour of such a symbolic interpretation. He saw palisades and trenches as emblematic of boundaries to places that were centres of community pride and prestige. This could explain Pā such as Ruahihi with its shallow ditch and bank and wide entrance seemingly providing little actual defence (McFadgen & Sheppard 1984, 36). The concept of Pā as a symbolic structure gained more momentum in later decades. Many archaeologists argued that the role of Pā was as a visible marker of a community that conveyed status, mana, and wealth; visibility therefore was emphasised (e.g. Barber 1996, Burridge 1995, Law 2000, Leach 2003, Sutton et al. 2003, Anderson 2009). A notable example of this is at Pouerua Pā (Figure 2.1) where flanking terraces and elongated roofed pits combined with a sculpted skyline created an impressive view of the Pā (Law 2000). Similar cases have been noted for many other Pā (e.g. Selwyn 1844).

These symbolic structures may also have been employed for religious purposes. Captain James Cook’s interpreter Tupaia identified Pā as places of worship (Beaglehole 1968, 191), and Mihaljevic (1973, 178-179) has argued that some Pā, particularly those of the South Island, served as temples rather than having a defensive function. There structures therefore may have been designed to reinforce

Figure 2.1: Aerial photograph of Pouerua Cone Pā (P5/195). Courtesy of Kevin Jones
these religious aspects. For example, some have argued that the elevation of posts in Pā reinforces the separation of Papatūānuku the Earth Mother and Ranginui the Sky Father in the Māori origin myth (Grey 1885, Sahlins 1985, Sutton 1991). Davidson (1984, 185) on the other hand argues that while Pā may have some symbolic function, this should not be overemphasised at the expense of the defensive function.

Stories of battles taking place at various Pā gleaned from ethnographic studies and oral history accounts have been used by archaeologists as evidence of the Pā’s defensive function (e.g. Best 1927, 2001, Brailsford 1981, Firth 1927, Smith 1984, Vayda 1960, Walton 2001). However, such studies can be flawed in that these ethnographies were conducted some time after European colonisation and oral histories were translated and edited by Europeans, so their reliability may be called into question. In addition to this ethnographical and oral history evidence, archaeological excavations of Pā have provided material evidence of warfare at Pā sites in the form of weapon and bone fragments just outside Pā palisades believed to be indicative of hand-to-hand fighting (e.g. Bellwood 1971a, b, 1978, Brailsford 1981, McKinlay 1971). However, as stated above, the number of Pā that have been excavated is very small and such examples cannot be considered representative of all Pā.

In spite of this evidence, Lilburn (1985) has argued that the settlement function of Pā is just as important as the defensive function. Indeed, a large proportion of Pā studies have been framed as settlement pattern analyses. A common theme in such studies is whether Pā were permanently or temporarily occupied or, alternatively, whether they should be treated as defended settlements or refuges/citadels.

Before the advent of professional archaeology in New Zealand, there was general consensus among scholars that Pā were defended villages (Gorbey 1970, 26). As archaeology became more scientific and systematic, this issue became more contentious. Kennedy (1969) argued for sustained occupation of Pā while allowing for some seasonal mobility. This view was supported by Bellwood (1971a) who drew on the perceived “insufficient evidence” of undefended permanent villages as confirmation that Pā must have been consistently occupied. However, many later excavations (e.g. Leach & Leach 1979, Sutton 1990) have found evidence of kainga – undefended sedentary villages. Some archaeologists (e.g. Fox 1976, 1983) have drawn on accounts from early British explorers and missionaries of housing and occupation on Pā sites as evidence of permanent occupation. However, Groube
(1965) argues that such accounts are misleading as the presence of Europeans would have forced the Māori into Pā. Instead he favours the citadel hypothesis where Pā are used irregularly and only in times of political tension. This hypothesis has been supported by many archaeologists (e.g. Buist 1964, Ballara 1979, Sutton et al. 2003, Orchiston 1979, Philips & Campbell 2004) with Sutton et al. (2003) stating that evidence of permanent occupation in excavated fortified sites is lacking. An exception to this rule that has been argued by several archaeologists is the Bay of Islands where evidence suggests Pā may have been permanently occupied (Gorbey 1970, Groube 1965).

In addition to these two functions Pā may have also served as fortified stores that were not occupied. Taniwha Pā (Figure 2.2) is a notable example of a Pā that is widely agreed to have served as an unoccupied foodstore. This is evidenced by the large number of pits (encompassing over 60% of flat land area) and lack of structures (Law & Green 1972). However, Law and Green do not rule out the possibility that this Pā was also used as a site of retreat. Gorbey (1970) suggests that such fortifications were built with the knowledge that conflict were possible but they were utilised for storage until such conflict arose. He builds on this argument by utilising the example of Pukearuhe Pā, which was initially a fortified store and later became occupied (Jones 1994, 52-53,57-58,145,148-149).

![Figure 2.2: Aerial photograph of Taniwha Pā (N58/1). Courtesy of Kevin Jones.](image)
Davidson (1984, 1987) argues that whether Pā functioned as settlements, citadels, or food stores varied regionally and individually, with the function of many Pā changing several times and being dependent on local variables such as terrain, economic need, social relations, iwi association, ideology, and time. This can be observed at Pouerua Cone Pā (P5/195) (Figure 2.1), which displayed significant change in function over time with fortifications only being functional for some of its history and the use of the area shifting from houses and occupation to what Sutton et al. (2003) argue was a more ceremonial use.

The international literature on fortifications and enclosures stresses that these sites were likely built for more than just one function (Neustupny 2006, Parkinson & Duffy 2007). Additionally, as Best (1927) has argued, Pā function varied and the specific meaning and use of any one Pā should be considered within its own context.

### 2.3 Pā Distribution

While few studies have gone into great depth on the distribution of Pā throughout New Zealand, from as early as the 1960s many archaeologists have described and sought to explain the general patterns observed. In the 1940s geographers partitioned New Zealand into three regions based on their suitability for sustaining pre-European populations (Cumberland 1949, Lewthwaite 1949). By descending suitability these regions were **Iwaitini**, the northern North Island, **Waenganui**, the central and southern North Island and very northern South Island, and **Te Wāhi Pounamu**, the remainder of the South Island (Figure 2.3). These geographers asserted that fortifications were less common in Waenganui and rare in Te Wāhi Pounamu due to the utilisation of forests for protection that obviated the need for Pā. Therefore these regions could potentially also reflect the distribution of Pā in Aotearoa. Many archaeologists have attested to the usefulness of these three regions for describing the general distribution of Pā within New Zealand (e.g. Davidson 1984, Groube 1965, Gorbey 1970, Vayda 1960, Walton 2002). However, Groube (1965) warns that there is considerable variation within these regions so the fact that a given location is situated in one of these regions alone is not enough to indicate the density of Pā in that area.

Buist (1964) has suggested that Pā locations are correlated with areas that have ready access to rich and desired food resources such as ridge tops for hunting, waterbodies for fishing, and arable land for agriculture. Most archaeologists have
Figure 2.3: The three zones described by geographers as reflecting areas with different suitability for sustaining the pre-European human population: Iwitini (yellow), Waenganus (blue), Te Wahi Pounamu (green).


When analysing the role of the Pā, Higham (1967) noted that the evidence of the time suggested a riverine and coastal distribution. Fox (1976) likewise
noted this association while Simmons (1969) and Cassels (1972) suggested that this apparent relationship was due to a reliance on kaimoana for protein as the Māori had no domestic animals (aside from the dog), which necessitated a close proximity to major water bodies.

Buist (1964) suggested that the most important food resource determining the location of Pā was agriculture, with soil quality being an important factor. In his study of 104 Taranaki Pā he demonstrated that the majority of Pā were located on the best soil for agriculture (although as noted above this study employed a fairly strict definition of Pā). This was supported by a nation-wide study conducted by Les Groube (1970) that suggested that 98% of Pā were located in the “zone of agriculture.” This pattern led Anderson (2009) to assert that “Pā are clearly a manifestation of agricultural development in some way.” However, this association has not been explored in all regions. Additionally, Davidson (1984, 1987) has cautioned against inferring too much from this relationship, claiming that Pā were superimposed on a pre-existing settlement pattern, and that any correlation with agricultural land is likely actually a correlation with areas of high population.

The first extensive study of the national distribution of Pā was carried out by Kenneth Gorbey (1970, 1971) and utilised the locations of 3100 Pā. This analysis involved comparing the gross environmental potential of an area, factoring in climate, soils, vegetation and coastal resources, to the density of Pā in that area in order to discern if there were an association between the two. His study concluded that while most Pā fell within Iwitini it did not identify a definitive correlation between easy environments and Pā density as had been anticipated. While most gaps in the distribution of Pā corresponded to rugged terrain or infertile soils, there was an unexpectedly large number of areas with very high environmental potential where Pā were under-represented or absent. The great variation in Pā density between areas of similar environmental potential lead Gorbey to conclude that “a postulate framed in environmental terms cannot explain the distribution of Pā within New Zealand” (Gorbey 1970, 126). Alternatively, he suggested that the distribution of Pā may be explained by relatively early settlement and population growth in areas where environmental potential was high. This study may be flawed due to the small number of Pā used - less than half of the number recorded today. The results therefore may be artificially produced by poor coverage by his sample of the true distribution of Pā. In addition, the methods used to determine correlation between Pā and easy environments are, by today’s standards, fairly unsophisticated, possibly leading to false results.
More recently, Leathwick (2000) has carried out a preliminary analysis of the distribution of Pā and pit sites across New Zealand using additive logistic regression. This method predicts the probability of observing a Pā or pit site in a given location based on environmental correlates. The resulting model gave high predicted probabilities of occurrence in high isolation, warm summer-dry locations, particularly in northern and eastern New Zealand, and in close proximity to major water bodies. This study provides a picture of the distribution of Pā but the use of “pseudo-absence points” calls the validity of these results into question. This is a randomly selected sample of points of an equivalent size to the Pā (“presence”) dataset drawn from a 1 km grid across New Zealand. This approach may introduce bias as the absence points are not at the same resolution as the Pā locations (1km vs. 100m), which may create artificial patterns, and because there is no guarantee that Pā truly are absent from these locations; they may have just not been recorded.

These examples demonstrate that spatial analysis is a valuable tool for investigating Pā, with many gaps in the literature still to be filled.

### 2.3.1 Social Organisation

While some archaeologists have drawn inferences about the social organisation of the Māori people from the spatial distribution of Pā, this has largely been limited to intra-site or regional studies. Nigel Prickett (1982, 1983) demonstrated in his study of the Pā of the Omata, Oakura, Tataraimaka, and Okato districts (Taranaki) that the proximity of Pā to one another at both the local and regional levels reflected the relationships between social groups and their lands. Geoffrey Irwin (1985) also noted the significance of the proximity of Pā to one another at Pouto (Northland), suggesting that Pā were located in relation to one another in contrast to open sites, which were generally located in relation to economic resources. This difference led Irwin to suggest that Pā must have a role beyond that of fortified habitation. The interpretation he gave for this observed pattern was that Pā represent the hapū level of organisation with most competition and conflict occurring between hapū groups, and that the spatial hierarchy of Pā reflected a political hierarchy. The study of hilltop Pā overlooking entrances to the Waihou river valley system undertaken by Phillips (2000) produced similar results. Phillips demonstrated that smaller, secondary Pā were located a short distance from larger ones, interpreting this spatial hierarchy as a manifestation of a political hierarchy.
These studies demonstrate that spatial organisation and social organisation are interconnected and bring into question what patterns may be observed in analyses on a larger scale. They also demonstrate that the relationships between Pā locations are important factors to consider when analysing Pā distribution as well as the association with resources that the other studies previously discussed have been focused on.

2.4 Conclusions

This review has demonstrated that there are still many holes in the literature on Pā and their distribution that remain to be filled. Three specific questions have emerged as a result of this review that will frame the following statistical analyses:

1. **What is the density distribution and clustering pattern of Pā sites in New Zealand?**

   While many archaeologists have described the general distribution of Pā, this has usually been based upon visual inspection of “distribution maps” (plots of raw point data), which can only provide the most basic information about the distribution of Pā in New Zealand and is subject to bias. There is clearly a need for more sophisticated methods that can provide more insight into this distribution and introduce more objectivity to our analysis. For this, statistical methods will be employed. Of particular interest will be whether these analyses provide any evidence for the validity of the three geographical regions described in this review as reasonable indicators of Pā density. If so this would provide some evidence for the association with agriculture that is so commonly posited by archaeologists. Additionally, the theory put forward by Davidson (1984, 1987) that Pā may be superimposed on a pre-existing settlement pattern should be explored by testing for evidence of an association between Pā density and population density, accounting for any relationship that exists when describing the density of Pā. The studies reviewed in this chapter have also noted the importance of Pā clusters in studying pre-European Māori social organisation, specifically clusters consisting of a large Pā surrounded by smaller satellite Pā. The locations of Pā sites across Aotearoa will need to be analysed in order to determine whether they form clusters, and if so the degree of clustering that exists, beyond the limited regions tested in the studies discussed.

2. **What are the dominant variables governing the density and clustering of Pā sites across New Zealand?**
Many archaeologists have identified food resources as one of the major factors governing Pā distribution based on the proximity of Pā sites to these resources. This has also been based largely on visual inspection of site locations and some use of simple statistical tests (Leathwick (2000) being the only major exception). The association between arable land, water bodies (kaimoana), and hunting spots should be measured using more sophisticated statistical methods in order to objectively determine whether there is evidence that this association truly does exist and to what extent. Again, the possible confounding effect of the overall population distribution will need to be considered. The statistical methods employed will need to be able to isolate any association between Pā and resources as distinct from any existing association between the pre-European population. Of interest will be whether the perceived association between Pā and food resources is in fact actually an association between the population distribution in general and these resources, with the Pā distribution being governed at least in part by the distribution of people. Additionally, competition for arable land has been proposed as a trigger for the instigation of Pā building. Therefore, measuring the degree of association between Pā and arable land would give an insight as to whether this hypothesis is likely to be true.

3. What are the implications of the identified patterns in terms of social organisation?

Of particular interest will be whether the results of our statistical analyses provide any new insights into pre-European Māori social organisation. Whether the models of social organisation put forward in the regional studies discussed above are supported by our findings on a national scale or whether new models need to be considered to explain the nation-wide pattern of Pā distribution will need to be explored.
Chapter 3

The Data

In order to answer the questions posed in the previous chapter we require reliable locational information for the Pā sites in New Zealand. For the following analyses, this information will be obtained from ArchSite - the online database of archaeological sites in New Zealand (NZAA 2009). This database consists of user-submitted archaeological site records and is operated by the New Zealand Archaeological Association (NZAA) in partnership with the Department of Conservation and Heritage New Zealand Pouhere Taonga.

3.1 A Brief History of Site Recording in NZ

In order to best utilise and understand this database, it is important to know where the information contained in it came from and how it was collected. The current site recording system in New Zealand, ArchSite, has its origins in the New Zealand Archaeological Association Site Recording Scheme. This was a paper-based site recording system established in 1958 with the aim of recording pre-European anomalies that had the potential to be investigated by archaeological methods in the future. This would later be expanded to include historic sites as well.

These paper site records consist of a site record form provided by the NZAA (see Figure 3.1) with fields that are filled in based on site visits, and any combination of site plans, section drawings, photographs, artefact drawings, and field notes. However, despite the existence of site recording handbooks that provided some guidance for recording sites (Walton 1999, Daniels et al. 1979)), there was no enforced standard for how these records were completed, with some including
extensive detail and others the bare minimum with only a location and site type given.

With regards to spatial information, in the early days of paper records site locations were given as imperial New Zealand Map Grid coordinates and later converted to the metric New Zealand Transverse Mercator. These coordinates ostensibly mark the south-east corner of a 100 x 100 metre square in which a site should be located. In reality however, due to errors in maps, the coordinate conversion process, and the difficulty of accurately pinpointing the location of remote sites in particular, this is often not the case. The consequence of this is that the error in these coordinates can often be greater than reported. More recently recorded sites have made use of handheld GPS technology to give precise coordinates rather than grid references, which are accurate to within ten metres.

Paper site records are organised by volunteer file-keepers in 18 different districts. A duplicate set of each file was also housed in Wellington in a collective central file, managed by the Central File-keeper.

With the introduction of the Resource Management Act 1991 and the Historic Places Act 1993 (replaced by the Heritage New Zealand Pouhere Taonga Act 2014) to protect heritage in New Zealand, the NZAA site records began to be utilised for statutory purposes, thus changing the nature of the site recording scheme. Whereas initially recorders were interested in potential archaeological sites, the site recording scheme now required accurate and reliable information on definite archaeological sites. For this reason, from 1999 to 2007 the NZAA carried out the Upgrade Project. The aim of this project was to update the quality of information on sites already recorded under the Site Recording Scheme. Existing records were reviewed and, if necessary, revisited in order to update and improve the corresponding information. This included checking and updating site locations, sometimes with GPS coordinates. However, this project was not carried out in all areas and some records from the Upgrade Project have yet to reach the central files.

In 2009 ArchSite moved archaeological site records online. Existing paper records were scanned and uploaded in bulk to ArchSite with the site type, coordinates and site 'ethnicity' (Māori or European) being added to searchable fields of that sites online record. Because information was digitally scanned directly from paper documents, this has resulted in some errors. These errors include some documents being attached to the wrong NZAA ID number, or missing completely.
### NEW ZEALAND ARCHAEOLOGICAL ASSOCIATION

#### SITE RECORD FORM (NZMS260)

<table>
<thead>
<tr>
<th>NZMS 260 map number</th>
<th>NZMS 260 map name</th>
<th>NZMS 260 map edition</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>NZAA METRIC SITE NUMBER</th>
<th>DATE VISITED</th>
<th>SITE TYPE</th>
<th>SITE NAME: MAORI</th>
<th>OTHER</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Grid Reference</th>
<th>Easting</th>
<th>Northing</th>
</tr>
</thead>
</table>

1. **Aids to relocation of site** *(attach a sketch map)*

2. **State of site and possible future damage**

3. **Description of site** *(Supply full details, history, local environment, references, sketches, etc. If extra sheets are attached, include a summary here)*

4. **Owner**
   - Address

5. **Tenant/Manager**
   - Address

6. **Nature of information** *(hearsay, brief or extended visit, etc.)*
   - Photographs *(reference numbers and where they are held)*
   - Aerial photographs *(reference numbers and clarity of site)*

7. **Reported by**
   - Address

8. **Filekeeper**
   - Date

7. **Key words**

8. **New Zealand Register of Archaeological Sites** *(for office use)*
   - NZHPT Site Field Code

<table>
<thead>
<tr>
<th>Latitude S</th>
<th>Longitude E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of site</td>
<td>Present condition &amp; future danger of destruction</td>
</tr>
<tr>
<td>Local environment today</td>
<td>Security code</td>
</tr>
<tr>
<td>Land classification</td>
<td>Local body</td>
</tr>
</tbody>
</table>

**Figure 3.1:** NZAA Site Record Form.
Additionally, some site type codes have been misread, resulting in some sites being entered into ArchSite with the wrong site type. At the time of writing, these issues were being addressed through a nation-wide audit of the paper records in each district file.

New site records and updates to existing site records can be made by any individual after acquiring an ArchSite account. This process involves filling a number of information fields and uploading any additional documents such as site plans and photographs. Once these are submitted they must be reviewed by the Central Filekeeper to ensure that they are of a high enough quality to be included in ArchSite and used for statutory purposes. Currently around 70% of site records cannot be approved in the first instance because additional information is needed from the site recorder. As a result of this, at any given time there exist around 1,000 sites in pending status.

### 3.2 Implications for the Following Analyses

The process that New Zealand’s recording of archaeological sites has gone through over time has given rise to several issues with the data that will need to be considered in this analysis:

- Sites that still have New Zealand Map Grid coordinates are, in theory, only accurate to 100m. In practice the error can be even larger than this for some sites. Sites that retain their original NZMG coordinates tend to fall in blocks marking the areas where the Upgrade Project was not carried out or data from it has not yet been entered into ArchSite. This may introduce some systematic bias to any statistical analyses of this point pattern if the error in site location is greater in some areas of New Zealand than others.

- Sites that are in the ‘pending status’ are not included in data exports so some sites will be missing from the following analyses.

- When the paper records were scanned some site type codes were misread resulting in some sites having incorrect site types in ArchSite and in the exported datasets. There is no way of knowing the exact number without going through every site record. However, this does appear to affect only a small number of site records.

- Because site recording has not been carried out systematically but rather has been undertaken by recorders in locations of their choosing, the absence
of a record for any given location does not necessarily indicate the absence of an archaeological site at that point but may indicate that the area has not been surveyed. This is more likely an issue for remote and difficult to access areas, and site types with poor above-ground visibility such as gardens. This could also introduce systematic bias to any analyses if these unrecorded sites fall in blocks as they may do if an entire area has not been surveyed.

- Because the site recording scheme has no strict guidelines for what sites should be classified as Pā or any other site type this is up to each individual recorder as noted by Schmidt (1996b) and discussed in Chapter 2. Therefore what sites are recorded as Pā or another site type will likely vary from site to site and possibly from region to region as individual archaeologists tend to record sites in a particular area.

- Because the Upgrade Project was not carried out in all areas and some information from the Upgrade has not yet reached the central records, some of the early potential sites still exist in the records that may not truly be archaeological sites. As stated above, the sites that lack upgraded information tend to fall in blocks, which, again may introduce systematic bias if certain areas have sites incorrectly classed as Pā, or sites in general.

Because the following spatial analyses are being undertaken at a national scale, this should reduce a lot of the error that these issues introduce to the point of being negligible. However, these factors and how they may affect any analyses undertaken henceforth will still need to be taken into consideration when selecting data and appropriate methods, and when interpreting any results.

### 3.3 Our Datasets - Filtering the Data

The following analyses will make use of spatial statistics methods. As we wish to analyse the distribution and clustering of Pā on a national scale, these will require point location data of the Pā sites in both islands of New Zealand. For completeness and to allow comparisons we will also require the locations of archaeological sites in general. The entire ArchSite database was exported in the form of a geodatabase (.gdb) on 05/05/2016. The attribute table for the sites layer in this geodatabase was then converted into an Excel spreadsheet in ArcGIS (ESRI 2016). This spreadsheet consists of a row for each site within ArchSite and twenty one corresponding columns for each ArchSite data field. Of relevance to this study are
the columns ‘NZTM_E’ and ‘NZTM_N’, which give the New Zealand Transverse Mercator easting and northing, respectively, of a given site, the ‘Island’ column, which indicates whether the site is located in the North or South Island, and the self-explanatory ‘SiteType’ column.

Before any statistical analyses can be undertaken, we need to extract and filter the relevant data from this table. We begin by splitting the data into two subsets. One subset consists of all those sites that have “Pā” as their site type, and the other consists of all those sites that do not have “Pā” as their site type, henceforth “non-Pā”. Both of these datasets were then filtered in order to make the data appropriate to answer the questions posed in this thesis and to help abate the issues discussed in Section 3.2.

**Filtering the Pā data:**

1. All post-contact/gunfighter Pā were removed. This involved removing all records with the “pa - gunfighter” SiteType and all sites where the Period field did not contain “Indigenous pre-1769”. While many archaeologists argue that the warfare that necessitated the construction of gunfighter Pā was a continuation of pre-European warfare rather than the result of the introduction of muskets (Ballara 2003), the very nature of musket warfare required that gunfighter Pā be built in different locations to traditional Pā, i.e., level ground as opposed to areas of steep topography. Additionally, evidence allows us to say with confidence that gunfighter Pā were constructed specifically for warfare and then immediately abandoned (Prickett 2002), which conflicts with the complex use histories for pre-European Pā discussed in Chapter 2. For these reasons it has been deemed inappropriate to analyse the spatial pattern of pre- and post-contact Pā together and the latter has thus been removed.

**Filtering the non-Pā data:**

1. Sites were removed if the Period field did not contain “Indigenous pre-1769” with the exception of sites where this field has been left blank. In this case these sites were left in at this stage. As we are only considering pre-European Pā sites our corresponding non-Pā dataset must like-wise only consist of pre-European sites in order to make the two datasets comparable in the following analyses. As discussed in Section 3.2, many sites have empty fields resulting from the shift from paper to electronic records. Therefore, many of the sites
that do not have a Period specified are likely to be pre-European. Thus, these sites are left in at this stage.

2. All sites with “Non-Maori” as their Ethnicity were removed. As above, non-Māori sites were removed to ensure that the non-Pā dataset was comparable to the Pā dataset in being pre-European. Again, many sites had a blank in this column. These were also left in the dataset at this stage.

3. Sites with the Site Type “Artefact Find” were removed. “Artefact Find” describes a location where an isolated artefact has been recorded. As such sites do not necessarily reflect human settlement or activity they are not appropriate for answering the questions posed in this thesis.
4. Sites with the Site Type “Traditional Site” were removed. “Traditional Site” refers to a site for which no physical evidence exists, but whose existence is assumed based on traditional accounts such as oral histories. As we cannot be sure that these sites exist, and in particular that they are recorded in the correct location, it would be inappropriate to include them in any analyses of site distribution.

5. All sites with a Site Type exclusive to the post-contact era were removed. As discussed above in Items 1 and 2, many sites did not have a Period or Ethnicity specified, making it impossible to filter out all post-contact sites using these columns. However, as noted in Section 3.2 all sites have a Site
Type selected. Thus, we are able to remove any remaining post-contact sites by removing all sites with exclusively post-contact site types.

Ideally, only sites contemporary with our Pā sites, i.e., from circa 1500 CE to the contact period, would be included. Early sites are common in the far south, and only in scattered coastal locations in the North Island. This is in contrast to later sites which are mostly in the North Island and display a shift toward inland locations. However, the dating information required to filter this data simply does not exist. This should not affect the following analyses significantly as these early sites should be far outnumbered by later sites that are contemporary with Pā, therefore these few outliers should not skew the distribution of non-Pā sites, particularly for the North Island.

These two datasets were subsequently converted to spatial point pattern objects in the statistical package R (R Core Team 2017), using the spatstat library (Baddeley et al. 2015). The point patterns for the North and South Islands are plotted in Figures 3.2, 3.3, and 3.4.
Chapter 4

Exploratory Analysis: Pā Density

The first question we have posed in this thesis concerns the distribution of Pā sites in the two major islands of New Zealand. As discussed in Chapter 2, in the past archaeologists have visually assessed the distribution and density of Pā by observing and plotting their locations (e.g. Higham 1967, Simmons 1969, Groube 1970, Cassels 1972, Fox 1976). This approach relies on subjective evaluation of visual patterns, and does not allow us to distinguish the causes of Pā distribution in particular from the causes of archaeological site distribution in general. We therefore desire some objective method of evaluating the density and distribution pattern of our Pā sites. For this, we turn to the field of spatial statistics.

4.1 Background: Spatial Statistics

Before we begin our analysis we must first consider some of the fundamental concepts of spatial statistics. These are necessary for the digestion of this chapter and also the formulation of the methods that will be employed to address our research questions.

4.1.1 Point Pattern

The term point pattern is used in spatial statistics to describe observed data that is made up of finite (i.e. countable) point locations in continuous space. These points will generally be contained within a known bounded region or window, denoted as W, which is a crucial component of many spatial analysis methods. This thesis will deal only with two-dimensional (2D) point patterns with polygonal windows. Examples of 2D archaeological point patterns are recorded find-spots of Anglo-Norman coins within the geographical border of the English mainland (Bevan
2012), or the centroids of Thai Bronze Age burials within the boundaries of the excavated area (Figure 4.1) (Smith et al. 2015).

![Figure 4.1: Map of excavated Phase Four Bronze Age burials at Ban Non Wat (left). Plot of the point pattern of centroids of Phase Four Bronze Age burials at Ban Non Wat (right).](image)

### 4.1.2 Point Process

The term point process is used to describe the underlying process that has generated an observed point pattern. This process may be a stochastic mechanism, function, or some other phenomenon, or combination of phenomena that controls the behaviour of the observed points. Any assumptions that we make about the underlying point process of a given point pattern will have a great impact on selecting appropriate methodologies for addressing problems and how we interpret the results of these methods. For archaeological point patterns, common underlying processes are often a combination of geographical/environmental limitations, human choices, or a random/stochastic component.

### 4.1.3 Intensity and Density

The intensity is one of the most fundamental properties of a point process, dictating the structure and appearance of the observed point pattern. However, in general the true intensity is not known so a great deal of spatial analyses involve estimating the intensity of a given point process based on these observed point
pattern. The intensity of a point process can be interpreted as the expected number of points per unit area when situated at an arbitrary location \( x \in \mathbb{W} \), and is denoted by \( \lambda \). The intensity can be expressed more formally as

\[
\lambda(x) = \lim_{|\delta x| \to 0} \left\{ \frac{\mathbb{E}[n(\delta x)]}{|\delta x|} \right\}; \quad x \in \mathbb{W}
\]  

(4.1)

where \( n(\mathbb{W}) \) is the number of points within the finite region \( \mathbb{W} \), \( \delta x \) is a small circular region centred at a location \( x \), and \( \mathbb{E}[n(\delta x)] \) is the expected number of points within that region. The constant or ‘overall’ intensity for an observed point pattern has the simple and intuitive estimator

\[
\hat{\lambda} = \frac{n(\mathbb{W})}{|\mathbb{W}|},
\]  

(4.2)

where \( |\mathbb{W}| \) is the area of the region. In practice the structure of an observed point pattern may be better described by a non-constant intensity, the estimation of which will be covered in the following section.

For some applications, it can often be more useful to think in terms of the density of a point pattern, \( f(x) \), rather than the intensity, \( \lambda \). The density and intensity are closely related due to the intensity having the property

\[
\mathbb{E}[n(\mathbb{W})] = \int_{\mathbb{W}} \lambda(x) dx,
\]  

(4.3)

i.e. the intensity integrates to the expected number of points falling within \( \mathbb{W} \).

Therefore, as densities by definition integrate to 1, we can see from Equation (4.1) that the continuous probability density \( f(x) \) of our point pattern can be found by simply scaling Equation (4.1) by \( n(\mathbb{W})^{-1} \). The density is more useful than the intensity when, for example, we wish to compare two point patterns with differing numbers of observations, because it avoids arbitrary scaling by the size of each dataset.

As stated above, the intensity (and related density) is one of the fundamental properties of a point process and the focus of many spatial analyses as it describes and importantly quantifies the structure and appearance of a point pattern. Sound estimation of the density of a given point pattern thus significantly reduces the subjectivity involved in the point location map approach previously employed by archaeologists. The true density of the point process that produced a given point pattern is generally unknown; this is certainly the case for the Pă location data.
One of the most popular and reliable methods of density estimation that does not require the true intensity to be known is kernel density estimation (KDE) (Rosenblatt 1956, Parzen 1962, Cocoullos 1966, Wand & Jones 1995).

4.2 Density Estimation: Method

Kernel smoothing refers to a nonparametric technique for estimating the density function of an observed point pattern that we cannot assume belongs to a common distribution family such as the normal distribution.

Consider the hypothetical 1-dimensional dataset \{7, 9, 12, 16, 17, 18, 19\}. These values are plotted on the \(x\)-axes of the plots in Figure 4.2 below. The kernel smoothed density estimate is calculated by assigning each observation in our dataset a weight defined by a particular kernel function \(K_h(z)\). These weights are ‘smoothed’ with respect to a chosen bandwidth \(h > 0\). The weights assigned to each observation are represented by the red curves centred over each point in Figure 4.2. For each plot a different bandwidth is employed to find the smoothed weights for each observation in order to illustrate how the bandwidth affects the kernel density estimation: the larger the bandwidth used, the more smoothed out the weights will be. The kernel density estimate for the dataset is then found by summing the weights at each location along the \(x\)-axis. The final estimate is represented by the blue line in each plot.

Over-smoothing by using too large of a bandwidth may result in losing some of the finer detail in the data as can be seen in the fourth plot where \(h = 3.5\). Under-smoothing by too small a bandwidth may give us a density that is too noisy, highlighting unimportant variation in the data as can be seen in the first plot where \(h = 1\). Selection of a “well-balanced” value of \(h\) is therefore imperative as we want to capture the general behaviour of the data.

This methodology is easily translated to two-dimensions by following the same procedure (Figure 4.3). Each point, is assigned a weight by a kernel function and this weight is smoothed out from this point with respect to a chosen bandwidth but now in two-dimensions. This smoothed weight is represented by the red disks (in reality the smoothed weight continues asymptotically and so does not have a defined boundary, but this disc indicates the main area of influence of that smoothed weight). As with one-dimensional smoothing, the kernel-smoothed density estimate is then found by adding the all of the weight at each location within the study area.
Once we have selected an appropriate kernel function $K_h(z)$ and bandwidth $h$ we can calculate the kernel-smoothed density estimate $\hat{f}_h(x)$ of our point pattern. Let $(x_1, x_2, ..., x_n)$ be our observations. The KDE at location $x$ is given by

$$\hat{f}_h(x) = \frac{1}{Nh^2} \sum_{i=1}^{N} K \left( \frac{x - x_i}{h} \right).$$

(4.4)

There are many choices for the kernel function that may be used to estimate $\hat{f}_h(x)$ for any given point pattern. For our observed point pattern of Pä site locations we will employ a Gaussian kernel (Equation 4.5) as is standard practice for geographical data (for an extensive discussion of kernel functions and their
selection see Wand & Jones 1995). This is given by

$$K(u) = (2\pi)^{-1} \exp \left( -\frac{1}{2} u^T u \right),$$  \hspace{1cm} (4.5)

where $u$ is the $2 \times 1$ coordinate vector.

For a rough initial estimate of the Pä density, the bandwidth $h$ is selected by using the rule-of-thumb of taking the smallest possible rectangle that encapsulates the window $\mathcal{W}$ of our point pattern and setting $h$ to be one eighth the longest side of this window (Baddeley & Turner 2005). By substituting these into Equation 4.4, we are able to calculate the Pä density estimates for each island. However, these estimates may be subject to what is known as boundary bias. This occurs when
observations fall outside of or close to the boundary of \( W \) (in our case the coastline of each island). If points fall outside of the window then they are not able to contribute to the kernel density estimate. If points fall close the boundary then some of the kernel weight assigned to them will fall outside of the window, as illustrated in Figure 4.3, and will therefore not contribute to the kernel density estimate. Our dataset of archaeological sites will naturally be affected by the former as the error in site locations, particularly for sites with NZMG coordinates, causes some site locations to fall beyond the New Zealand coastline. Additionally, some of the kernel weight assigned to points near the coast will fall outside of the window resulting in strong negative bias of the kernel density estimate at the edges of the window. This is a serious problem for our dataset as such a large proportion of Pā are located very close to the coastline.

In order to correct for this bias we apply *edge correction*. This is a process where the KDE is adjusted in order to reduce edge bias to an asymptotically negligible level (Diggle 1985, Kelsall & Diggle 1995, Marshall & Hazelton 2010). This correction is applied simply by rescaling the density estimates by division at each point \( x \in W \) by

\[
q_h(x) = \int_W \frac{1}{h^2} K \left( \frac{z-x}{h} \right) dz.
\]  

(4.6)

Equation 4.6 can be interpreted as the proportion of the kernel weight at location \( x \) that falls within \( W \). Therefore, more weight is given to points where the kernel falls partially outside the window, making up for the lost kernel weight and thereby providing an approximate cancellation of edge bias.

The edge-corrected kernel-smoothed density estimates are plotted as heat map-style images that use colour gradients to illustrate how the value of the density changes across the region. These are produced for each island and plotted in Figures 4.4 and 4.5 below. Note that these are calculated and plotted independently so shared colours between the two plots do not indicate the same density in each island.

### 4.3 Density Estimation: Results

From Figure 4.4 we can observe that the density of Pā is at its greatest in the north of the North Island (Northland) and that this density steadily decreases
as we move south, with the lowest density falling in the Wellington region. By comparing this density to the geographical zones in Figure 2.3, we can see a clear correlation between the area designated as Iwitini and the area where Pā density is comparatively high, with Pā density dropping close to zero at a point that matches reasonably closely the borders of the Iwitini and Waenganui regions. The observed pattern in the density of the South Island given in Figure 4.5 is similar. The density is strongest in the north-east of the South Island and demonstrates a steady but rapid decline as we move south-west. Again, the drop in density matches closely the borders of the Waenganui and Te Wāhi Pounamu zones shown in Figure 2.3.

As discussed in Chapter 2, it is commonly argued that Pā are associated
with agriculture in some way with many researchers arguing that Pā were built to protect arable land as population and, subsequently, competition for arable land grew. Our density estimate would appear to lend support to this argument. Pā are most dense in Iwitini, where agricultural conditions are generally best, than they are in Waenganui where the environment is considered more marginal. However, the usefulness of these results is affected by many of the issues previously explored. In Chapter 2 we noted that Davidson (1984) has argued that Pā may have been superimposed on a pre-existing settlement pattern. If this is the case we would expect the estimated density of Pā to reflect, at least in part, the density of the pre-European human population in New Zealand. The correspondence of Pā density with these three geographical zones supports a relationship with
population as researchers have suggested that around 80% of the pre-European Māori population lived in Iwitini, where our kernel density plots indicate Pā were most dense, 15% in Waenganui, and 5% in Te Wāhi Pounamu, where the density is lowest (Cumberland 1949, Lewthwaite 1949). This connection makes it difficult to extract any meaningful or interesting patterns from the density of Pā as it is likely to be, at least in part, dictated by population density. Instead, these plots may be heavily confounded with the background population distribution, which tells us nothing about Pā and their role in the pre-European settlement pattern.

In addition, as discussed in Section 3.2 Pā sites may be missing from this dataset due to destruction or non-discovery leading to systematic bias that may make any interpretations of these densities unreliable. Hodder & Orton (1976, 237) propose that the issues of survival and discovery bias may be alleviated by comparing the distribution of interest with another distribution. If one can assume that the site types making up two distributions have similar chances of survival and discovery then differences between these distributions will be informative.

One such method for comparing distributions is the density ratio, also known as the relative risk method, which was developed for geographical epidemiology. This method was produced as a means of coping with the confounding effects of population (Kelsall & Diggle 1995, Bithell 1990, 1991). Thus, it is an ideal tool for coping with both issues that we have identified. Because of its utility this method has been used in many other fields including archaeology (e.g. Bevan 2012, Bevan et al. 2013, Smith et al. 2015).

### 4.4 Density Ratio: Method

The density ratio method involves calculating the ratio between two densities, called the ‘case’ and ‘control’ densities (following conventional terminology used in epidemiology), in order to identify areas where the cases are significantly more or less dense compared to the controls. Therefore we must select an appropriate control distribution to compare to our Pā distribution (cases). We require a site type that has similar chances of survival and discovery to Pā for the reasons discussed above. Additionally, as we want to remove the confounding effects of population as much as possible we want to select a site type with a distribution that is similarly governed by population density.

The obvious choice is to use the locations of all other pre-European site types as our control point pattern. We can reasonably assume that these sites, which
we will henceforth refer to as ‘non-Pā’ sites, have similar rates of survival and
discovery to Pā across New Zealand, thus accounting for this potential source of
bias. Secondly, it is also reasonable to assume that the distribution of these sites is
correlated with population density as there is naturally more human activity, and
therefore more sites created, where there are more people. This is supported by
observing the plot of the density of non-Pā sites in the North and South Islands in
Figures 4.6 and 4.7, calculated using exactly the same method as for Pā outlined
in Section 4.2. We can observe that these sites are very dense in Iwitini and
relatively sparse in Te Wāhi Pounamu.

Therefore, by using non-Pā sites as our controls our question of interest shifts
from “where are Pā?” which we have demonstrated is problematic, to “where are
Pā more or less likely to occur compared to other pre-European site types?” This
makes interpreting the resulting density ratio very intuitive and will allow us to
make more reliable and more interesting interpretations based on our results than
we are able to get from a simple density estimate.

We assume that the Pā and non-Pā point patterns are realised random sam-
pleS from densities \(f\) and \(g\), respectively. These densities are estimated by finding
the two-dimensional kernel-smoothed density estimates (Diggle 1985), denoted as
\(\hat{f}_h\) and \(\hat{g}_h\), using Equation 4.4.

In order to calculate \(\hat{f}_h(x)\) and \(\hat{g}_h(x)\) for these point patterns we must first
select an appropriate kernel function \(K_h(z)\) and bandwidth \(h\). When calculating
geographical density ratios it is standard practice to employ a Gaussian kernel
(Equation 4.5) (Hazelton & Davies 2009, Davies & Hazelton 2010).

Next we need to select an appropriate bandwidth \(h\). So far we have only
discussed fixed bandwidth kernel density estimation where \(h\) is chosen to be a
constant. Alternatively we can choose what is known as an adaptive or variable
bandwidth approach. This is where a variable smoothing parameter is used for each
kernel. Adaptive bandwidths have been shown to be more appropriate for relative
risk estimation when the densities of interest are not uniform (Davies & Hazelton
2010). The distribution of Pā sites (and non-Pā sites) is strongly inhomogeneous,
with areas of high Pā density and areas where it is practically impossible for a
Pā to be located (rivers, lakes, mountains etc). For such point patterns a fixed
bandwidth KDE may fail to capture some of the finer detail in areas where the
density of observations is high when a large amount of smoothing is undertaken
in order to reduce excess noise in more sparsely populated areas. Conversely if
a small amount of smoothing is employed to better capture the detail in dense
areas then the resulting relative risk surface will likely have spurious bumps in areas with a lower density of points due to isolated points introducing too much noise into the KDE. In other words, when carrying out fixed bandwidth KDE it may be difficult to achieve an appropriate balance between capturing important detail in areas where Pā or other site types are dense, and reducing noise in areas where they are sparse. This results in a relative risk surface that is biased and/or difficult to interpret. Adaptive bandwidth KDE avoids this problem by ascribing a variable amount of smoothing that is inversely related to the density at any given area.

We thus adjust the kernel density formula slightly to become
Figure 4.7: Kernel-smoothed estimate of the probability density function of non-Pä sites in the South Island.

\[ \hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h(x_i)^2} K \left( \frac{x - x_i}{h(x_i)} \right) \]  

(4.7)

where \( h(.) \) is a function that controls the amount of smoothing. For this analysis we will use the popular variable bandwidth, which has been shown to result in improved properties compared to a fixed bandwidth (Abramson 1982). For completeness we will calculate and compare both fixed and adaptive kernel density estimates, henceforth distinguished using the subscripts \( f_{fix} \) and \( f_{ad} \).

For our estimation of fixed bandwidth KDEs there are more sophisticated jointly optimal estimators for the bandwidth but their validity has been called
into question. For simplicity we will employ the common oversmoothing bandwidth (Terrell 1990). Hazelton & Davies (2009) and Davies & Hazelton (2010) demonstrated that this simple bandwidth selection methodology performs well in practice when calculating density ratios. For an overview of bandwidth selection, and the relative risk function in general, see Davies et al. (2017).

The oversmoothing bandwidth is defined by

\[ h_{OS} = U \times \left[ \frac{10^4 \pi R(K)}{16n \Gamma(20)^8} \right]^{\frac{1}{6}}, \quad (4.8) \]

for the 2D case. \( U \) is a scalar estimate of the standard deviation of our data, and \( R(K) = \int K(u)^2 du \).

With our kernel density estimates \( \hat{f}_{fix}, \hat{g}_{fix}, \hat{f}_{ad} \) and \( \hat{g}_{ad} \), we are now able to evaluate the relative risk function \( r(x) \) and explore the relationship between the distributions of Pā and non-Pā sites. The spatial relative risk function is defined as

\[ r(x) = \frac{f(x)}{g(x)}; \quad x \in \mathbb{W}; \quad (4.9) \]

simply the ratio between the case and control densities. By convention, the natural log of the relative risk function is used in order to stabilise tail values (Kelsall & Diggle 1995), giving a revised formula of

\[ \rho(x) = \ln(r(x)) = \ln(f(x)) - \ln(g(x)); \quad x \in \mathbb{W}. \quad (4.10) \]

Therefore, the log-relative risk \( \rho(x) \) simply describes the size of the difference between the logs of the case and control densities at any given point in the region \( \mathbb{W} \).

The relative risk surface \( r \), and subsequently \( \rho \), is estimated by substituting the kernel density estimates of \( f \) and \( g \) found above into the relative risk function giving \( \hat{r}_h \) (and by extension \( \hat{\rho}_h \)). The estimated log relative risk function \( \hat{\rho}_h \) is then plotted as a heat map-style image in order to show how the value of the log relative risk changes across the window \( \mathbb{W} \), or island as it is in this case.

To aid in the interpretation of a log relative risk surface, tolerance contours (Hazelton & Davies 2009) are typically superimposed onto these heat-map plots. Tolerance contours are lines that distinguish statistically significant peaks and troughs from random noise in the log relative risk surface. In the context of our study, these contours identify areas where the density of Pā is significantly high or
low compared to the density of other sites. Statistical significance is determined by testing the natural null hypothesis of uniform risk, i.e., that the density of the cases (Pā) and the density of the controls (non-Pā) are equal at a given location, giving $\rho = 0$. P-values for this hypothesis test are obtained by employing a normal approximation to the sampling distribution of $\hat{\rho}_h$ (for full details we direct the reader to Hazelton & Davies 2009).

Setting the confidence level to a standard 95%, areas of the relative risk surface with a corresponding P-value of less than 5% are considered to be statistically significant and are marked on the relative risk map with a contour line. A solid contour line delineates statistically significant peaks in the relative risk surface while a broken contour line delineates statistically significant troughs.

4.5 Density Ratio: Results

The relative risk surfaces for the North and South Islands, using both fixed and adaptive bandwidths, are plotted below with tolerance contours superimposed to aid in our interpretation (Figures 4.8 to 4.11).

Because we are using the natural log of the relative risk (Equation 4.10), areas where the value of the risk surface is positive indicate a higher density of Pā than non-Pā and where the risk surface is negative, the opposite holds. This makes the log relative risk surface very intuitive to interpret.

From the two log-relative risk surfaces for the South Island we can observe that regardless of whether a fixed or adaptive bandwidth is employed, the resulting plots demonstrate very similar patterns. The density of Pā is significantly high compared to that of non-Pā in the north-east corner of the South Island and on Banks Peninsula, and significantly low in most of the bottom half of the South Island. The most notable difference between the two risk surfaces is that the adaptive relative risk identifies Banks Peninsula as having a significantly low density of Pā compared to non-Pā. Overall, both plots appear to show fairly similar patterns to the simple density estimate for the South Island. These plots may be considered a little too smooth. This is likely being driven by the small number of observations over such a large area. For this reason we will restrict attention to the North Island.

The two relative risk surfaces for Pā in the North Island on the other hand reveal a markedly different pattern when compared to the simple density estimate.
Chapter 4 Exploratory Analysis: Pā Density

Figure 4.8: Fixed-bandwidth relative risk surface for Pā sites compared with non-Pā sites in the North Island of New Zealand. Solid lines delineate areas where the relative risk is significantly high, and broken lines delineate areas where the relative risk is significantly low.

of Pā sites given in Figure 4.4. While the density of Pā shows a steady and consistent decrease from north to south across both islands, the relative risk surfaces for the North Island show a more complex pattern. The tolerance contours highlight that the relative risk is significantly low in the Wellington region where the density of Pā is also relatively low. However, the most surprising feature of these plots is that the relative risk of observing a Pā is significantly low in the Auckland region, revealing that while the density of Pā in this area is very high, as shown in Figure 4.4, there is actually a significantly low density of Pā when compared to the density of all other site types. Additionally, through the central North Island where the stand-alone density of Pā starts to become low, Pā actually have
4.6 Conclusions

Understanding the distribution and density of Pā is an important step towards being able to understand their role in pre-European Māori social organisation. Despite this, research into the distribution of Pā has largely been limited to small-scale regional studies. Archaeological studies investigating the distribution of Pā on a national scale have largely relied on visual inspection of location maps, which is generally subjective. In this chapter we have employed kernel density estimation...
in order to curb this subjectivity to a certain extent. Based upon the inspection of our KDE plots, our results appear to support the conclusions of previous researchers: Pā are most dense in the northern North Island, with a gradual decrease in density moving southwards. South Island Pā are most dense in the north-east corner of the island, including Banks peninsula.

However, we question the validity and usefulness of a stand-alone density approach to quantifying the distribution of Pā due to the strong association between the distribution of Pā and the distribution of the pre-European Māori population. By instead utilising density ratios that compare the density of Pā to the density of other pre-European archaeological site types we have accounted for the
confounding effects of population and potential bias from missing data.

The estimated relative risk of Pā is significantly low in the north of the North Island where the density of Pā is high and significantly high in the central North Island where the density is relatively low. One could argue that this pattern is being driven by the high prevalence of midden sites in the non-Pā dataset. Middens are highly visible sites that are recorded in large numbers; the ‘midden/oven’ site type is outnumbered only by ‘pit/terrace’ sites. These sites tend strongly towards the coastlines of New Zealand and therefore could conceivably be driving the observed pattern. However by estimating the log relative risk of Pā against the six most prevalent non-Pā site types individually (Figure 4.12) (the remaining site
types had fewer than 100 observations each) we can observe that the relative risk pattern is very similar for all site types except for cave/rock shelter sites (this may be due to the fact that the locations of this site type are driven more by availability than any other site type); the relative risk is generally low through the Northland/Auckland region and high through the central North Island. Therefore the relative risk estimate using all non-Pā sites does not appear to be biased by any one site type.

The marked differences between the estimated relative risk surface and the Pā KDE surface indicate that, while Pā density is governed to some extent by population, there are areas where there are significantly more or fewer Pā than would be expected based on the distribution of archaeological sites in general. These areas of significant difference should therefore be very informative and need to be examined more closely.

Returning to the suggestion that Pā have a strong association with arable land, it is of note that the relative risk of Pā is significantly low in the region with the greatest agricultural potential (Cumberland 1949, Lewthwaite 1949) and significantly high through the central North Island where the environment is more marginal. This does not necessarily discount the notion of association but could suggest that the relationship is instead inverse; a possible explanation for this observed pattern is that Pā were built in relatively great numbers to defend arable land where it was more scarce and therefore more precious whereas it was not as necessary to build Pā to defend agricultural land in the north where it was so abundant. However, as we discussed in Chapter 2, some researchers argue that there was enough arable land to sustain the pre-European population, with extensive tracts in the Waikato, which makes up part of our region with a significantly high relative risk of Pā, remaining undeveloped. The relationship between Pā distribution and agriculture/arable land therefore needs to be explored further in order to ascertain if there is evidence of an association between agriculture/arable land and areas where the relative risk of Pā is high or low.

This analysis has highlighted how a straightforward density approach to analysing Pā fails to capture the subtleties of the Pā distribution that may provide real insight into Pā and their role in the pre-European social organisation that studies have hitherto left unexplored. By exploring differences between the significant regions identified by our relative risk estimates and by testing factors that may contribute to changes in the relative risk surface, we should gain more insight.
Figure 4.12: Fixed-bandwidth relative risk surfaces for Pā sites compared with six non-Pā site types in the North Island of New Zealand. Solid lines delineate areas where the relative risk is significantly high, and broken lines delineate areas where the relative risk is significantly low.
For the following analyses we will look only at the North Island of New Zealand. The small number of Pā observations and potential temporal bias caused by the large number of early sites as discussed in Chapter 3 make any further statistical analyses of the South Island difficult.
Chapter 5

Exploratory Analysis: Pā Clustering

Another important feature of the Pā distribution is the relationships between individual Pā sites. In particular we are interested in whether Pā “attract” each other. In other words, can we conjecture that Pā tended to be built in close proximity to other Pā? Such a phenomenon would result in an aggregated point pattern. In spatial statistics aggregation refers to point patterns where there are significantly more points within a given distance $r$ of some point $x$ than there would be if the distribution of said point pattern were uniform. These points therefore form clusters. The opposite case is called repulsion and describes when there are significantly fewer sites within a given distance $r$ of some point $x$ than would be expected if the point pattern were uniform. It is possible for a given point pattern to exhibit both aggregation and repulsion at different distances. For example sites may form clusters at small distances but these clusters may in turn repel each other at larger distances. If we are able to identify aggregation and/or repulsion of our Pā sites, and at what distances this occurs, this will provide more insight into the pattern of distribution of our Pā and consequently the pre-European Māori settlement pattern.

Archaeologists have generally identified clusters of Pā through visual inspection of maps (e.g Buist 1964, Prickett 1982, 1983). However as previously discussed, this can be very subjective and patterns are sometimes inferred from completely random points (Hodder & Orton 1976, 4). Some archaeological studies have used statistical methods to identify clusters, such as Irwin (1985) who used a nearest neighbour index (NNI) approach to detect clustering. This involves calculating the index $I$ using $I = \bar{D}/\left(0.5 \times \sqrt{\frac{|W|}{n}}\right)$ where $\bar{D}$ is the mean observed distance between each point and its nearest neighbour. Standard cut off values are then used to determine whether the NNI indicates a clustered, random, or regular
point pattern. Nearest neighbour methods are fairly popular in archaeology, how-
however they can be flawed as they are ‘short-sighted.’ Because they only consider the
nearest neighbour of each point, they are not suitable for describing the structure
of a point pattern at larger distances and can give misleading results in some cases
(Beyer et al. 1999). This is especially of concern for more complex point patterns,
such as that of Pā.

Therefore we desire some statistical method that is capable of measuring
aggregation and repulsion at a nation-wide scale.

When seeking to describe correlation between any two points in a point pat-
tern spatial statisticians turn to second-order summary statistics. ‘Second-order’
refers to statistics that describe relationships between points in a point pattern,
thus making them ideal for exploring correlation. One of the most commonly used
and recommended of these statistics is the pair-correlation function, which, as the
name implies, measures the strength of the correlation (or association) between
pairs of points for a given point pattern. While the estimation of this statistic
depends on the satisfaction of certain assumptions it is considered the most infor-
mative of the second-order summary statistics employed in spatial statistics as its
results are very readily interpretable (Illian et al. 2008, 214-218).

5.1 Background: Spatial Statistics

Again, there are some spatial statistics concepts that are necessary to understand
and implement the pair-correlation function.

5.1.1 Complete Spatial Randomness, Homogeneity and
Heterogeneity

Often one of the first questions a researcher will ask themselves when faced with
point process data is ‘is it random?’ Or, put more formally, ‘is there evidence of
some underlying structure to our point pattern that would not likely exist under
Complete Spatial Randomness (CSR)?’ If we have $N$ points within the bounded
region $W$ then CSR implies that the observed points are an independent sample
from the uniform distribution on $W$. Uniform describes a distribution where the
probability of observing a point at a given location $x$ is equal for all $x \in W$. In
spatial statistics the term homogeneous is used to describe the underlying process
of such a point pattern. An inhomogeneous, also known as heterogeneous, point
process will have an intensity/density that varies across space. In other words, a heterogeneous point process is not uniform.

The simplest example of CSR is the *homogeneous Poisson point process*, where our points follow a Poisson distribution with mean \( \lambda |\mathcal{W}| \) (Diggle 2003), where \( \lambda \) denotes the intensity as defined in Section 4.1.3.

Plot A in Figure 5.1 below gives an example of a point pattern generated using a homogeneous Poisson point process. A simple visual inspection supports the lack of any clear structure to the point pattern, suggesting it to be uniform across space. Plot B was generated using an inhomogeneous Poisson point process. The presence of small clusters are clear evidence against CSR. In general, however, it is not possible to assess whether a point pattern exhibits CSR through visual inspection alone and statistical methods must be employed, such as the pair-correlation function.

![Figure 5.1: Two hypothetical examples of 2D point patterns. Plot A was generated using a homogeneous Poisson point process. Plot B was generated using an inhomogeneous Poisson point process.](image)

### 5.2 Pair-Correlation Function: Method

The PCF function \( g(r) \), where \( r \) is the radius of a disc centred at a randomly selected point \( x \), can be mathematically defined as:

\[
g(r) = \frac{\lambda_2(r)}{\lambda^2}. \tag{5.1}
\]
In this formula $\lambda$ denotes the intensity of the point pattern (found using Equation 4.1), while $\lambda_2$ denotes the second-order intensity given by

$$
\lambda_2(x, y) = \lim_{|\delta x|, |\delta y| \to 0} \left\{ \frac{\mathbb{E}[n(\delta x)n(\delta y)]}{|\delta x||\delta y|} \right\}; \quad x, y \in W; \quad x \neq y. \quad (5.2)
$$

By comparing the form of the second-order intensity $\lambda_2$ to that of the intensity $\lambda$ introduced in Section 4.1.3 it should become clear that the second-order intensity can be loosely interpreted as the expected product of counts per unit areas at any two arbitrary locations $x, y \in W$.

As the true values of $\lambda$ and $\lambda_2$ are not known $g(r)$ must be estimated. The standard estimator introduced in Stoyan & Stoyan (1994, 284–285) employs the kernel smoothing technique described in Chapter 4 giving

$$
\hat{g}(r) = \frac{\hat{\lambda}_2(r)}{\hat{\lambda}^2} = \frac{1}{2\pi \hat{\lambda}^2 r |W|} \sum_{u \neq v} K_h(||u - v|| - r) w_R(u, v); \quad 0 < r \leq r_{max}, (5.3)
$$

where, following standard practice, $K_h(x)$ is the univariate Epanechnikov kernel (Equation 5.4) and $w_R(u, v)$ is the Ripley edge correction (Stoyan & Stoyan 1994). $\hat{\lambda}$ is found using the estimator given in Equation 4.2. Therefore, while the kernel smoothing employed in Chapter 4 was applied to our observations, the PCF carries out a univariate smooth on the pair wise distances between those observations.

$$
K_E(u) = \frac{3}{4} h^{-1} \left( 1 - \frac{u^2}{h^2} \right) 1(-h \leq u \leq h) \quad (5.4)
$$

The pair correlation function thereby describes the correlation in the point pattern by comparing the expected number of points per unit area for the whole region to the expected number of points per unit area at a given distance $r$ apart. This allows us to see how the local intensity of points at given distances varies and differs from the global intensity, thus allowing us to identify aggregation and repulsion. Because the PCF considers a range of distances $r$ and looks at the number of points within that distance, it is not subject to the same issues as nearest neighbour approaches.
In order to assess whether the actual estimated values of \( g(r) \) indicate clustering or repulsion we compare our result to the theoretical PCF calculated under complete spatial randomness (CSR). Under the assumption of CSR, \( \lambda_2(x, y) = \lambda(x)\lambda(y) = \lambda^2 \), due to independence. Therefore Equation 5.1 will give a theoretical value of 1 for the CSR case. If the estimated value of \( g(r) \) for our observed Pā data exceeds 1 for a given distance \( r \), this indicates that the second-order intensity \( \lambda_2 \) is greater than the squared intensity \( \lambda^2 \), meaning that there are more points per unit area at distances of \( r \) than expected based on the global intensity of the region. This therefore provides evidence of clustering in the Pā point pattern at the given distance \( r \). If the \( \hat{g}(r) \) is less than 1 then the opposite is true and we have evidence of repulsion of Pā sites.

The PCF is estimated for a range of \( r \) values and plotted in Figure 5.2 with a horizontal line at 1 indicating the theoretical PCF value under CSR. To further aid interpretation, a 95% tolerance envelope is also included in the plot. This is a region in which we are 95% confident that the estimated PCF will fall if the observed point pattern is indeed a case of complete spatial randomness. In other words, the tolerance envelope represents a range of values in which we could reasonably expect our estimated PCF to fall if there were in fact no clustering or repulsion of Pā sites. This region is estimated by generating 39 iterated point patterns from the Poisson (CSR) density within the North island window \( \mathbb{W} \), with number of points equal to that of the original dataset, and calculating \( \hat{g} \) of each of these point patterns. The boundaries of the tolerance envelope are then given by calculating the central 95% quantiles of the resulting distribution.

### 5.3 Pair-Correlation Function: Results

The plot of the pair correlation function calculated for the point pattern of Pā sites in the North Island is given in Figure 5.2. The solid black line represents the value of the PCF for our observed point pattern at different distances, \( r \). The broken red line represents the theoretical value of the PCF for a Poisson point pattern (1∀\( r \)), and the grey band about the red line represents the 95% tolerance envelope for the theoretical PCF as described above.

As the PCF for the Pā point pattern sits entirely outside of the tolerance envelope for all values of \( r \) this indicates that we have very strong evidence that our data do not follow a Poisson (CSR) distribution and the location of a given Pā does influence the location of other Pā to some extent. Additionally, as the
PCF value sits entirely above the 95% tolerance envelope for all distances, this indicates that we have more Pā than expected within any given distance of a Pā. This means that the presence of a Pā at a given location tends to attract other Pā. In other words, Pā tend to be members of large clusters.

More insight may be gained into the clustering and distribution pattern of Pā by estimating the pair-correlation function for the Pā contained within the significant regions from our log-relative risk estimate. We will focus on the most notable regions from the relative risk surface in Figure 4.8: the significantly low northern region, and the significantly high central region. The resulting estimates are given in Figure 5.3.

Based on these plots it appears that there is still strong evidence of clustering in each region. However in contrast to the clustering pattern of the North Island as a whole, the clusters have a limit to their size as evidenced by the PCF dropping within the confidence band. These clusters are of different sizes for each region: around 15 km in diameter in the northern region and 50 km in diameter in the central region (Illian et al. 2008). Also of note is that in the central region the PCF immediately drops below and stays below 1 indicating that there are less Pā than expected beyond a distance of 50 km of another Pā. This evidence indicates that the clusters of Pā repel each other at this and larger distances. The PCF for the northern region also shows evidence of repulsion as it too dips below.
Figure 5.3: Pair-correlation functions for the northern and central relative risk regions with close up views. The solid black line represents the value of the PCF for the observed point pattern, the broken red line represents the theoretical value of the PCF under CSR, and the grey band represents the 95% tolerance envelope for the theoretical PCF.

1. However, the PCF then re-enters and sits within the confidence band. This means that the number of Pā within a distance of 50-55 km of each other is not significantly different from what would be expected under a Poisson distribution. This indicates that the clusters in this northern region are randomly distributed at these distances.

The pair correlation function as we have estimated it assumes that all of the heterogeneity in the Pā point pattern is due entirely to correlation between these sites. This is likely not a sensible assumption for this point pattern as other
factors can reasonably be expected to have influenced the deviation from a uniform distribution that we see besides clustering of Pā. Two obvious contenders are the fact that some areas are very remote and inaccessible and so Pā are less likely to be built in these locations (there was likely little demand for a Pā atop Mount Cook, for example); the other is the effect of the human population distribution as discussed in Chapter 4. If these and other factors are contributing to the heterogeneity of the Pā point pattern then the pair-correlation function will not be accurately describing the level of association (if any) between Pā sites. Fortunately there is a relatively simple solution to this problem: we can estimate and remove the heterogeneity in our point pattern, allowing us to estimate any remaining correlation between sites. This is carried out using the inhomogeneous PCF, $g_{inhom}(r)$, the estimation of which is similar to that for $\hat{g}(r)$ (Baddeley et al. 2000).

5.4 Inhomogeneous Pair Correlation Function: Method

The inhomogeneous pair-correlation function is given by

$$\hat{g}_{inhom}(r) = \frac{1}{2\pi r |W|} \sum_{u,v \in \mathcal{X}} K_h(||u - v|| - r) \frac{w_R(u,v)}{\hat{\lambda}(u)\hat{\lambda}(v)}$$

(5.5)

Here, $\hat{\lambda}(\cdot)$ is the estimate of non-constant intensity, i.e., our measure of heterogeneity, which is proportional to the kernel smoothed density estimate $\hat{f}_h(x)$ calculated using Equation 4.4.

We also need to select an appropriate bandwidth for our KDE. As stated previously, the bandwidth determines the amount of smoothing in a density estimate so we need to proceed with some caution in our selection. If the bandwidth used is too large then our KDE will be too smooth and thus not adequately capture and remove the heterogeneity in our point pattern meaning our inhomogeneous PCF estimate will not be reliable; any clustering detected may still be being driven by other sources of heterogeneity. If the bandwidth used is too small then the resulting KDE will be too noisy and overestimate the amount of heterogeneity to be removed, meaning we may not be able to detect any residual clustering. In this case it is difficult to say exactly how smooth our intensity estimate needs to be in
order to capture and remove the deterministic heterogeneity in our point pattern. Therefore a range of bandwidths are utilised so that the effects of removing different amounts of heterogeneity on the PCF can be explored. The oversmoothing bandwidth, $h_{OS}$ (Equation 4.8), utilised for our density ratios in Chapter 4 is used as a guide with bandwidths of half the oversmoothing bandwidth and twice the oversmoothing bandwidth also being employed. Again, 95% tolerance envelopes are generated using the same method described above to aid in our interpretation. The resulting inhomogeneous pair-correlation functions and tolerance envelopes are plotted below in Figures 5.4, 5.5, and 5.6.

![Inhomogeneous pair-correlation function for Pā in the North Island.](image)

**Figure 5.4:** Inhomogeneous pair-correlation function for Pā in the North Island (left). The bandwidth used is given by twice the oversmoothing bandwidth. A close up view of the same inhomogeneous pair-correlation function (right). The solid black line represents the value of the PCF for the observed point pattern, the broken red line represents the theoretical value of the PCF under CSR, and the grey band represents the 95% tolerance envelope.

### 5.5 Inhomogeneous Pair Correlation Function: Results

In Figures 5.4, 5.5, and 5.6 we can observe that the inhomogeneous PCF initially sits entirely above the grey tolerance band, indicating that we have strong evidence of clustering at these smaller distances regardless of what bandwidth is used to estimate the density of the Pā point pattern. The key difference in these plots is the values of $r$ for which the inhomogeneous PCF is significant. As the bandwidth
becomes smaller, giving us a less smooth intensity that removes more of the heterogeneity in our point pattern, the distances $r$ over which Pā exhibit aggregation decrease. For the homogeneous PCF calculated above we saw clustering at all distances of $r$. As more heterogeneity is captured by our kernel intensity estimate, this distance decreases to $r \approx 21\text{km}$, $r \approx 12\text{km}$, and $r \approx 7\text{km}$, respectively. As with the homogeneous PCF estimate, for all three cases, the PCF drops to below 1 for greater values of $r$ indicating that these Pā clusters tend to repel each other at greater distances.

Therefore, even when using the smallest bandwidth given by $\frac{h_{\text{OS}}}{2}$ to remove the largest amount of heterogeneity, the Pā point pattern still exhibits strong evidence of aggregation, albeit at much smaller distances. This indicates that some of the perceived attraction of Pā exhibited in the homogeneous PCF (Figure 5.2) at larger distances was actually due to heterogeneity in the point pattern brought about by other factors and not by clustering. However, all three plots clearly demonstrate that regardless of the level of smoothing employed we still have strong evidence of local clusters of Pā even after accounting for the underlying heterogeneity in their point pattern. Therefore, even in the presence of other factors influencing the distribution of Pā we still see strong statistical evidence of clustering of Pā sites.
We again compare the inhomogeneous PCF estimates of the central and northern relative risk regions of significance to identify differences in their clustering patterns. These are plotted in Figure 5.7 for the $h_{OS}$ bandwidth.

Once other sources of heterogeneity have been taken into account, the clustering patterns of these two regions are somewhat different. Both regions still exhibit strong evidence of clustering, but now these clusters appear to be about the same size at 10 km in diameter in both regions. This is just a small decrease in size for the northern Pā clusters but quite a large decrease for the central Pā from 50 km in diameter. We still see evidence that the clusters in the central region repel each other at all remaining distances as the inhomogeneous PCF value sits below 1 for these distances, so our interpretation does not change significantly. Our interpretation of the northern region, on the other hand, does change. As opposed to the PCF plot above where the PCF went to zero indicating that the clusters had a random arrangement, the inhomogeneous PCF exceeds 1 at distances of 50-55 km. While we must be cautious about interpreting the PCF at greater distances as it becomes less reliable, this does provide some that we have a second level of clustering in this northern region. The small clusters of Pā appear to form larger clusters of clusters at greater distances.
Chapter 5 Exploratory Analysis: Pā Clustering

5.6 Conclusions

The clustering pattern of Pā is another important feature of the distribution of Pā that contributes to our understanding of their role in the pre-European Māori settlement pattern. By employing the pair correlation function, we have demonstrated that Pā across the North Island do exhibit evidence of clustering. The PCF does not drop to 1 at any distance indicating that all Pā tend to be members of one big cluster. This is not a surprising result; the pair-correlation function assumes that all of the heterogeneity in a point pattern is due to clustering. A cursory examination of the Pā point pattern strongly suggests that it is not homogeneous.
with Pā concentrated in the north and along the coasts. When we account for deterministic heterogeneity from other sources by using the inhomogeneous pair-correlation function, we still see strong evidence of residual clustering, but we also now see evidence of repulsion at larger distances. This indicates that the apparent degree of clustering of Pā is exaggerated by other sources of heterogeneity, the two major contributors to which are likely geography and population distribution. These other sources of heterogeneity also blur the boundaries between clusters, completely obscuring the presence of repulsion in the Pā point pattern, and potentially between clusters.

By comparing the clustering patterns of the northern and central regions of significance from the relative risk estimate of the North Island we have identified further differences between these regions. While the central region displays some evidence of clusters that repel each other, the pattern in the northern region appears to be of clusters of clusters.

Our results suggest that the clusters of Pā that exist are likely not purely the result of factors such as population density or geography, but rather there is some social factor driving pre-European Māori to build Pā in groups. The differences in the clustering patterns of the northern and central regions therefore suggest regional differences in the social organisation of pre-European Māori.
Chapter 6

Modelling the Distribution of Pā

With a greater understanding of how the density of Pā varies across New Zealand, particularly in relation to other site types, the next logical question to ask is what factors are driving this variability? The few studies that have attempted to model the distribution of Pā have done so by analysing how the probability of observing a Pā at a given location is affected by environmental correlates at that location, or by testing the strength of the association between Pā and certain environmental correlates (e.g. Gorbey 1970, 1971, Leathwick 2000). Refer back to Chapter 2, Section 2.3 for details on these analyses. The reliability of such analyses is imperilled should the distribution of Pā be confounded by the pre-European population distribution, because Pā distribution is then simply an index of the broader pattern of human activity in different environments. Indeed, the results of these studies appear to be consistent with such a scenario. For example, the results of Leathwick (2000) match closely what we know about the distribution of pre-European Māori, with higher predicted probabilities of Pā in the north and near the coasts (Figure 6.1).

An alternative approach that will be developed and employed in this chapter is to model where the density of Pā is significantly high or low compared to the density of other pre-European site types. In other words, we will attempt to model our relative risk surface in terms of a collection of potentially important predictor variables.

6.1 Background: Statistical Modelling

The modelling approach developed in this chapter makes use of existing model fitting procedures that must be understood before proceeding any further.
Figure 6.1: Results of Leathwick (2000).
6.1.1 Generalised Least Squares

Generalised least squares is a method of estimating the unknown parameters of a linear regression model when the model residuals are correlated. This method involves first defining our response variable by placing all of our observation values into an $n \times 1$ vector of the form

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

where $n$ is the number of observations or responses. The values for our $k$ predictors are placed in an $n \times k$ design matrix of the form

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{k1} \\ 1 & x_{12} & \ldots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \ldots & x_{kn} \end{pmatrix}.$$

The column of 1s here represents our model constant/intercept. The remaining columns correspond to each of our $k$ predictors, with each row corresponding to one of our $n$ observations. Therefore, each cell $X_{ij}$ in our design matrix gives the value of the $i$th predictor for the $j$th observation.

Let us also define $\beta$ as a vector of the $k$ unknown regression coefficients $\beta_k$ corresponding to each of our $k$ predictors that we seek to estimate:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}.$$

The model we are trying to fit can therefore be expressed in matrix form as $Y = X\beta + \epsilon$:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{k1} \\ 1 & x_{12} & \ldots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \ldots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$
where $\epsilon_i$ is the residual error associated with the $i$th observation $y_i$. Generalised least squares differs from a simple regression approach by accounting for the correlation present between these residuals by incorporating a variance-covariance matrix $\Omega$ into the model fitting framework. This matrix consists of $n$ rows and $n$ columns. Each entry $\Omega_{ij}$ in this covariance matrix gives the covariance between the residual errors $\epsilon_i$ and $\epsilon_j$, thereby describing the correlation structure of our observations. $\text{Cov}[\epsilon_i, \epsilon_i]$ is equal to the variance of $\epsilon_i$, therefore the diagonal of the covariance matrix is made up of the variance of each residual, denoted as $\sigma_i^2$.

$$\Omega = \begin{pmatrix}
\sigma_1^2 & \text{Cov}[\epsilon_2, \epsilon_1] & \ldots & \text{Cov}[\epsilon_n, \epsilon_1] \\
\text{Cov}[\epsilon_1, \epsilon_2] & \sigma_2^2 & \ldots & \text{Cov}[\epsilon_n, \epsilon_2] \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}[\epsilon_1, \epsilon_n] & \text{Cov}[\epsilon_2, \epsilon_n] & \ldots & \sigma_n^2
\end{pmatrix}$$

By employing this variance-covariance matrix, the regression coefficients can then estimated by evaluating the equation $\hat{\beta} = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y$.

### 6.2 Geostatistical Modelling: Method

The proposed strategy is to set the value of the relative risk surface at each pixel as our response variable and then regress these values on our chosen predictors. This could be done using the straightforward linear regression approach but this methodology implies our observations (i.e., our relative risk pixel values) and subsequently our residuals are independent. This is clearly inappropriate: neighbouring pixels will tend to be highly correlated due to the smoothing process used in kernel density estimation. We must therefore account for this correlation, and we thereby move into the realm of geostatistics. Geostatistics is similar to spatial statistics in that it deals with spatially referenced data. Unlike spatial point process statistics however, it is specifically concerned with spatially correlated data, as opposed to being interested in the point locations/pattern itself. A good practical overview can be found in Bivand et al. (2013).

Consider the following simple example dataset: say we estimate the relative risk for two randomly generated inhomogeneous point patterns within a unit square window (Figure 6.2) using the same methods described in Chapter 4. A fixed oversmoothing bandwidth (Equation 4.8) and Gaussian kernel are employed, and the relative risk function estimated on a $64 \times 64$ grid of pixels (Figure 6.3). Under our proposed approach this gives a response variable with 4,096 observations.
Next we require an artificial spatial covariate to regress our risk surface against. To best illustrate this method we will define a covariate that is strongly associated with the risk surface. Let us create a pixel surface where the value of each pixel \( l_i \) is given by multiplying the value of each pixel \( y_i \) in the relative risk surface by 2 and then adding a randomly generated value \( v \) between 0 and 0.1 from the uniform distribution (Equation 6.1). We thereby define our covariate as the value of this pixel surface at a given location.

\[
l_i = 2y_i + v \quad v \sim U(0, 0.1)
\]  

Next we seek a parametric model to describe the correlation structure of our risk surface. This is found by employing variogram analysis, a geostatistical method for characterising spatial correlation. The first step in variogram analysis is to regress our response variable (relative risk pixel value) against our predictor variable \( l \) using ordinary least squares model fitting. As previously stated, the resulting residuals of such a model fitting approach will be correlated. We then estimate the semivariance \( g(h) \) of these residuals. The semivariance is a statistic that measures the variability between spatial points at a given distance or lag, \( h \), thus indicating the degree of correlation between pixels in our relative risk surface. Smaller values of \( g(h) \) indicate a high level of correlation, while larger values of \( g(h) \) indicate the opposite. Our notation is as follows:

- \( u \): A vector of spatial coordinates (the locations of our pixels)
Figure 6.3: Estimated relative risk for the two randomly generated point patterns within a unit square window.

- $z(u)$: The variable under consideration as a function of $u$
- $h$: The scalar lag or distance between any two locations
- $N(h)$: The number of pairs of observations separated by a lag of $h$

The semivariance $g(h)$ is given by

$$g(h) = \frac{1}{2N(h)} \sum_{a=1}^{N(h)} [z(u_a + h) - z(u_a)]^2$$  (6.2)
The resulting estimate $\hat{g}$ is given in Figure 6.4 as a variogram. The variogram plots the estimated semivariance against increasing lags $h$, illustrating how the estimated variability between the values of $z$ changes as the distance between those points increases. From this plot we can see that the semivariance between points increases and therefore the correlation between points decreases as the distance or lag increases, as is typical of spatially indexed data.

For this toy dataset and the Pā data below, the Matérn covariance function is fitted to the variogram using ordinary least squares (for full details on this model see Schlather 1999, 4). This function is defined in terms of the parameters $\kappa$, $\sigma^2$ and $\phi$, the estimates of which are given in Table 6.1.

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$\sigma^2$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.81444</td>
<td>0.15310</td>
</tr>
</tbody>
</table>

Table 6.1: Correlation parameters for our example dataset estimated using variogram modelling.

The fitted parametric model is plotted against the variogram in Figure 6.4 in order to visually assess its fit.

We can thus use this model to populate a covariance matrix for use in generalised least squares model fitting. This covariance matrix is made up of 4,096 rows and 4,096 columns where each row and each column corresponds to one pixel in the relative risk surface, and each cell therefore corresponds to a pair of points. The estimated Matérn model is then used to evaluate the covariance $\hat{V}_{ij}$ between each of these pairs of points using the distance $h$ between them thus populating the covariance matrix.

There is one last element of our data to consider before we can model it. While the relative risk surface is continuous with values at all locations within our specified window, the two point patterns from which the relative risk was calculated are not. We need to account for the fact that different amounts of data are available at each pixel location. This can be done with a reasonably simple adjustment to our covariance estimates.

Let $p_i$ be the value of the $i$th pixel of the pooled kernel density estimate of our two example point patterns. The pooled density is estimated by employing a Gaussian kernel and fixed oversmoothing bandwidth. Also let $a$ be the area of the $i$th pixel and $n$ be total pooled number of observations from our example point patterns. We define a weight $w_i$ at each pixel $y_i$

$$w_i = np_i a. \quad (6.3)$$
We then adjust our initial covariance estimates $\tilde{V}_{ij}$ by using the formula

$$V_{ij} = \frac{\tilde{V}_{ij}}{\sqrt{w_i w_j}}. \tag{6.4}$$

This adjusted covariance matrix is then plugged in to the generalised least squares equation procedure to fit a model to our relative risk surface that accounts for the pixel-to-pixel correlation. This model takes the form

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 l \tag{6.5}$$

The estimated parameters resulting from the generalised least squares are given in Table 6.2 along with the corresponding standard error and confidence
Estimate | Standard error | Confidence interval
--- | --- | ---
$\beta_0$ | -0.00252 | 1.89742 | -3.11429, 3.10924
$\beta_1$ | 0.49995 | 0.11000 | 0.31956, 0.68035

Table 6.2: Parameter estimates for our example dataset following generalised least squares.

The fitted model is given in Equation 6.6.

$$\hat{y} = 0.49995l - 0.00252.$$  \hspace{1cm} (6.6)

From the parameter estimate of 0.49995 for $l$ we can see that this method has provided a reasonable estimate of the relationship between our synthetic covariate and toy the risk surface.

Naturally the values of the pixels in the relative risk surface will change with different bandwidths. This thesis presents the first use of this particular method so there is no standard bandwidth or guide to selecting an appropriate bandwidth. Therefore we will calculate the relative risk using a range of bandwidths. The over smoothing bandwidth $h_{OS}$ is again used as a guide and our range of bandwidths is found by dividing the $h_{OS}$ by values from 1 to 4 at intervals of 0.5. This thus gives us a range of relative risk surfaces from the very smooth surface given by smoothing with respect to $h_{OS}$ to the relatively noisier surface given by smoothing with respect to $\frac{h_{OS}}{4}$.

### 6.3 Covariate Selection

The first step in our modelling procedure is to select appropriate covariates to use to model the relative risk of Pā compared to non-Pā sites. For this analysis we are interested in testing factors that archaeologists have commonly argued have a strong geographical association with Pā in order to explore whether these factors can account for areas where Pā are more or less dense than other site types. As discussed in Chapter 2 the distribution of Pā is most commonly linked to three factors; water bodies, ridge tops and arable land. Many studies at both a regional and national scale have concluded that Pā tend to be located on or near these three features (e.g. Buist 1964, Allen 1994, 1996, 2006, Anderson & McGlone 1992, Cassels 1972, Bellwood 1978, Duff 1967, Furey 2006, Groube 1970, Kirch 2000, Leathwick 2000, McGlone et al. 1994, Vayda 1960, Walton 2001, 2006). Water
bodies and ridges were both important locations for sourcing food; water bodies through fishing, eeling, and collecting shellfish, and ridges for hunting (Buist 1964). Fresh water bodies were also important of course as a source of drinking water. Having these natural features nearby would therefore have been important if Pā were being used as permanent settlements and possibly even more important if Pā were being used as refuges due to the threat of violence meaning people could not venture far for these important resources.

It is often argued that the most important role of Pā was to protect agricultural lands. Researchers cite the perceived association between Pā and good soil as well as storage pits as evidence of this role (Buist 1964, Groube 1970, Brailsford 1981, Barber 2004, Anderson 2009). A strong spatial association between Pā and areas suitable for growing Māori cultigens could therefore support this hypothesis. Alternatively such a pattern could also occur as a result of people wanting to grow their crops near their Pā.

The data employed to quantify these factors were sourced from the Land Resource Information Systems (LRIS) Portal (Landcare 2017) and Land Information New Zealand (LINZ) Data Service (LINZ 2017). The LRIS portal is a repository of authoritative New Zealand science datasets and information and is hosted by the informatics team at Landcare Research. The LINZ Data Service contains information on land titles, geodetic and cadastral survey systems, topographic information and hydrographic information.

### 6.3.1 Water Bodies

Detailed geographical data on New Zealand’s water bodies can be accessed via the LINZ data service in the form of topographical maps. Topographical maps of water bodies are divided into three categories; lakes, swamps, and rivers. In each case the different water body types are recorded as vector polygons at a scale of 1:50,000 for all the islands of New Zealand. This provides us with very fine detail data on the locations and extents of water bodies in New Zealand, making them ideal for use in this spatial analysis.

As we are interested in water bodies in general rather than having separate water body categories, these three polygon shapefiles were combined into a single layer. This layer was then cropped by our window of interest, i.e., the North Island coastline used in the preceding analyses. This therefore gives us a layer describing the location and extent of fresh water bodies in the North Island of New Zealand.
Next, as our response variable in this analysis is the pixel values of the log-relative risk surface, our predictor variables must similarly consist of values at the same locations as these pixels. Therefore the water body layer must be quantified in some way so that a value can be assigned to each of these pixel locations. One option would be to use the distance from the location of each relative risk pixel to the nearest water body. However due to the computational complexity of this modelling approach we are forced to lower the resolution of the relative risk surface to 128. At this lower resolution, the distance from any pixel to the nearest water body would be zero, so this approach would not be appropriate. Even if we were able to run the analysis at a higher resolution this approach would not be able to account for the differences in size of water bodies or points that have few or many water bodies nearby. We could restrict our analysis to only “major” water bodies as did Leathwick (2000) in his analysis, however this leads to the difficulty of deciding some cut-off for which water bodies would be considered “major”. Such a cut-off could only be arbitrary. Additionally this approach would give weight to locations with a large water body some distance away while ignoring locations with several smaller water bodies close by. Such a distinction may not be useful or informative.

We therefore opt for a novel approach that accounts for these factors. We have created a new layer with pixels matching the size and location of the relative risk pixels. We populate this layer with the percentage of area within the extent of each pixel that is covered by water based on our water body polygons (Figure 6.5). This new layer therefore does not discriminate based on water body size or number but instead looks at the amount of water in an area as a whole.

This leaves one major water body still unaccounted for; the ocean. This cannot be treated in the same manner as the other water body types as it sits outside of the region of our analysis. In addition, as it is not used for drinking water it may not be appropriate to consider together with fresh water as a single variable. Instead we take the distance from the centre of each relative risk pixel to the nearest point on the coastline, again using the same North Island coastline we used in the preceding analyses, giving us a ‘distance from the coast’ predictor variable.

6.3.2 Ridges

Ridges are a somewhat more difficult feature to quantify than water bodies. There is no readily accessible data that gives the locations of ridges in New Zealand.
Additionally, if such a dataset did exist it would be difficult to make it workable at the resolution this analysis requires. Instead we utilise slope as a proxy for ridge locations. If a ridge is the point where two slopes meet and Pā are associated with ridges, then there should also be some association between sloped areas and Pā. As slope is a continuous variable we will therefore use the degree of slope as our covariate.

Slope data for New Zealand can be accessed through the LRIS portal. This data comes in the form of a raster layer with a 25 metre resolution where the value of each pixel is the slope at that location in degrees. This resolution is too high for our analysis and thus needs to be reduced to match the resolution of the relative risk surface. We do this by employing the bilinear interpolation method.

Figure 6.5: Percentage of pixel area covered by freshwater bodies.
This method determines the new value of a pixel by taking the average of the 4 nearest pixels, weighted by distance. The resulting pixel image (Figure 6.6) therefore gives us the average slope at the locations of each of the pixels in the relative risk surface.

![Figure 6.6: Average slope in degrees.](image)

### 6.3.3 Agriculture

Quantifying agriculture in a meaningful way is not straightforward. Agricultural sites are difficult to identify based on surface surveys except where there have been major modifications such as stone rows or trenches. Even where excavation has been carried out, garden soils are not easily identified as large areas were able to be gardened without modifying the soil (Furey 2006). Therefore while agricultural
sites are recorded in ArchSite, attempting to determine how well the recorded sites represent the true distribution of gardens in New Zealand may be problematic. As it is generally where unusual or specialised methods were employed that are found and recorded, these sites may even provide a misleading picture of the distribution of gardens in New Zealand. Pit sites are sometimes used as a proxy for gardens following the assumption that pits were built near gardens in order to store the food grown there. These sites are also much more easily identified through surface surveys than gardens are and so the data may be more reliable, at least insofar as representing the distribution of pit sites in New Zealand. Again however, it is difficult to judge how well these sites, which are also recorded in ArchSite, represent the distribution of agriculture in pre-European New Zealand.

For these reasons we are not able to define a predictor variable that describes agricultural site locations directly. We can instead however employ environmental correlates that describe the agricultural potential of a given area or its suitability for growing Māori cultigens.

At the time of the first European explorations of New Zealand there were six introduced Polynesian cultigens being grown. These were kumara, taro, yam, gourd, ti pore (cabbage tree), and aute (paper mulberry) (Barber 2004). The most important and extensively grown was the kumara, likely because it was tolerant to the widest range of environmental conditions and matured faster. While each of these plants have different conditions for successful growth, they all depend on the same general environmental phenomena, if to different extents. Research from archaeologists indicates that the most important physical conditions for the successful growth of these crops were sunshine, temperature, and moisture (Leach 1984, Jones 1994, Furey 2006).

All of the crops listed naturally need sunlight to grow. As these plants evolved in and adapted to tropical Polynesian Island conditions, they grow best with higher levels of sunlight. The tropical origins of these plants also determined the range of temperatures in which they are able to grow, with all of these species thriving in warmer temperatures where they are not threatened by frost (Leach 1984, Jones 1994, Furey 2006). Experiments of modern tubers have found that tissue death occurs at temperatures below 5°C (Leach 1984).

Māori root crops also typically do not flourish with wet feet, with the exception of the taro. Overly wet soil, especially when combined with low temperatures, causes root rot in these crops (Leach 1984, Jones 1994, Furey 2006). Therefore, porous soil with good drainage is more suited to these crops as it prevents the build
up of excess moisture and therefore the roots from rotting. Porous soil also has the additional benefits of being easier to work with the wooden tools available to pre-European Māori and warming more quickly in the spring (Leach 1984, Furey 2006). Taro, the only crop that does prefer wet soil is best suited to swamps or other waterlogged areas. As we have already created a land wetness variable this has already been taken into consideration.

The data for all three variables are taken from the LRIS portal and are in the form of raster layers at a 25 m resolution. The sunshine data comes from the mean annual solar radiation raster, temperature from the mean annual temperature raster, and soil drainage from the soil drainage raster. The temperature and solar radiation raster layers are made up of continuous data so it is appropriate to
reduce the resolution in the same manner as used above for the slope raster layer (Figures 6.7 and 6.8). Soil drainage on the other hand is categorical with five classes of drainage ranging from Very Poor (1) through to Good/Well Drained (5). Because this layer is categorical it is not appropriate to reduce the resolution by averaging pixel values as was done for the other layers. Instead a new layer with pixels matching the size and locations of the relative risk pixels is laid over the soil drainage layer and each cell of this grid is populated with the most common drainage class within its area. As large areas tend to have only one drainage class, this method of reducing the resolution does not result in a significant loss of information (Figure 6.9).

All six parameters are summarised in Table 6.3 below.
All of these variables were used in Leathwick’s (2000) spatial analysis of Pā, with the exception of our differing description of water bodies. This analysis found all of these predictors to be significant factors contributing to the probability of observing a Pā in a given location. It will therefore be very interesting to see whether these variables can account for changes in the relative risk of Pā across space.
Table 6.3: Summary of covariates to be used in geostatistical modelling of relative risk.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Proportion of pixel area covered by fresh water</td>
<td>0–1</td>
<td>0.002</td>
</tr>
<tr>
<td>Temp</td>
<td>Mean annual temperature in °C × 10</td>
<td>7–160</td>
<td>125</td>
</tr>
<tr>
<td>Coast</td>
<td>Distance from pixel to nearest point on coast</td>
<td>0–1.292</td>
<td>0.286</td>
</tr>
<tr>
<td>Slope</td>
<td>Mean slope in degrees (°)</td>
<td>0–45</td>
<td>8</td>
</tr>
<tr>
<td>Soil</td>
<td>Most common soil drainage class</td>
<td>1–5</td>
<td>5</td>
</tr>
<tr>
<td>Sol</td>
<td>Mean annual solar radiation in MJ/m²/day × 10</td>
<td>136–154</td>
<td>146</td>
</tr>
</tbody>
</table>

6.4 Results - Single Covariates

We initially explore each predictor variable in isolation in order to ascertain whether any of these environmental correlates individually contribute to a significant spatial change in the distribution of Pä compared to non-Pä. The variogram modelling approach described above is employed to populate covariance matrices for use in fitting each of the six single covariate models tested via generalised least squares. The relative risk surface used in this analysis is estimated on a 128 × 128 grid of pixels. Due to the shape of the North Island not every pixel is assigned a relative risk value. These empty pixels are removed from the analysis meaning our response variable has 4,672 observations. The covariance matrix therefore consists of 4,672 rows and columns with each cell giving the covariance between relative risk pixels \( y_i \) and \( y_j \). Generalised least squares is then used to fit models with the structure

\[
\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i
\]  

(6.7)

Because soil is a categorical variable with five levels, the model has the form

\[
\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 d_1 + \hat{\beta}_2 d_2 + \hat{\beta}_3 d_3 + \hat{\beta}_4 d_4
\]  

(6.8)

where \( d_i \) are the dummy variables corresponding to the different levels of Soil. Soil = 1 is taken as the reference level. The resulting parameter estimates and their corresponding 95% intervals are given in Figures 6.10 and 6.11 below.

It is reassuring to note that the parameter estimates do not change significantly when different bandwidths are employed as all of the confidence intervals for each covariate overlap except those for the distance from coast parameter. However, these confidence intervals do still overlap for neighbouring estimates. The major difference we see with different bandwidths is that the estimated standard
error changes, altering the width of our confidence intervals. The narrowest intervals and therefore most precise parameter estimates are given when the relative risk estimated using the oversmoothing bandwidth is modelled. Additionally, the standard errors tend to increase as the bandwidth becomes smaller for all covariates. This is to be expected as there is more noise in the estimated relative risk for smaller bandwidths.

By looking more closely at these confidence intervals we can see that distance from the coast is the only covariate with confidence intervals that do not contain zero at $h < h_{OS}$. This indicates that we have sufficient evidence to suggest that distance from the coast does have some effect on the probability of observing a $\text{P}\ddot{a}$.
compared to a non-Pã site. As the parameter estimates for distance from coast are all positive, this indicates that as the distance from the coast increases, the relative risk of observing a Pã compared to a non-Pã site also increases.

Taking the parameter estimates from $h_{OS \frac{1}{1.5}}$, as this had the smallest standard error and is the most conservative estimate, the equation for the log relative risk in terms of distance from the coast in kilometres ($x$) is

$$\hat{\rho} = 0.003x - 1.0528. \quad (6.9)$$

Therefore, for every 1 km increase in distance from coast, the log relative risk of observing a Pã compared to a non-Pã increases by 0.003. When smaller
bandwidths are used, giving less smooth relative risk surfaces this effect increases, with a maximum increase in relative risk of 0.025 per kilometre.

There is no evidence to suggest that any of the other parameters tested have an effect of the relative risk, individually.

This process was repeated using the squared parameter variables (excluding soil because it is categorical and therefore squared values are not meaningful), with little change in the results. Again, distance from coast was the only parameter that exhibited any significance at any of the bandwidths tested.

### 6.5 Results - Multiple Covariates

While only one of the covariates tested has an effect on the relative risk surface individually, others may have some effect when combined. We will employ backwards selection where we fit the full model and then remove non-significant parameters in turn until we arrive at a model with only significant parameters. Significance is determined with a standard $t$-test and using a standard significance threshold of 5%. The results from fitting the models to the relative risk surface estimated using the $h_{OS}$ bandwidth are given here.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>0.006348</td>
<td>0.815108</td>
</tr>
<tr>
<td>Temp</td>
<td>0.000181</td>
<td>0.517764</td>
</tr>
<tr>
<td>Coast</td>
<td>0.000003</td>
<td>0.000055</td>
</tr>
<tr>
<td>Slope</td>
<td>0.000033</td>
<td>0.878000</td>
</tr>
<tr>
<td>Sol</td>
<td>-0.003227</td>
<td>1.638134</td>
</tr>
<tr>
<td>Soil2</td>
<td>0.000197</td>
<td>0.981081</td>
</tr>
<tr>
<td>Soil3</td>
<td>-0.000077</td>
<td>1.078458</td>
</tr>
<tr>
<td>Soil4</td>
<td>-0.002080</td>
<td>1.205662</td>
</tr>
<tr>
<td>Soil5</td>
<td>-0.002213</td>
<td>1.231404</td>
</tr>
</tbody>
</table>

*Table 6.4: Parameter estimates and corresponding P-values for the geostatistical model containing all six parameters.*

Table 6.4 gives the parameter estimates and P-values (to 5 decimal places) for each variable resulting from fitting the model with all six covariates. Observing these P-values we can see that distance from the coast is the only parameter with a P-value less than 0.05 and therefore is the only parameter that is significant in the presence of all other parameters. We remove the least significant parameter, soil drainage, and proceed to fit the model with the other five parameters.
Continuing to follow this procedure and fitting models with progressively fewer covariates, distance from the coast remains the only significant parameter in each model fitted. All variables are removed in turn until we are left with the single variable model containing distance only from coast. This process was repeated for the other bandwidths to the same result: distance from the coasts the only variable to exhibit significance in any of the multiple covariate models fitted. Each variable is removed in turn until only distance from the coast remains, for every bandwidth tested.

Interactions between pairs of variables were also tested in order to determine whether two of our environmental covariates combined had some influence on changes in the relative risk of observing \( \text{P} \). Each interaction, along with its corresponding main effects, was tested separately. Interactions involving soil were excluded as these models were too complex to fit with current computational limits. Multiple interactions were not tested in a single model for the same reason.

The resulting estimates and p-values of the interaction terms from each of these models, using the relative risk calculated using \( \frac{h_{0.5}}{h_{0.5}} \), are given in Table 6.5 below.

From this table we can see that none of these interaction terms are significant as their P-values are well above the standard significance threshold of 0.05. The same interaction models are fitted using the other bandwidths to the same results: none of the interaction terms are significant.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sol:Coast</td>
<td>-0.001173</td>
<td>1</td>
</tr>
<tr>
<td>Sol:Water</td>
<td>-0.001173</td>
<td>1</td>
</tr>
<tr>
<td>Temp:Water</td>
<td>(-1.0226 \times 10^{-4})</td>
<td>1</td>
</tr>
<tr>
<td>Temp:Coast</td>
<td>(3.8605 \times 10^{-9})</td>
<td>0.753404</td>
</tr>
<tr>
<td>Slope:Water</td>
<td>(-3.6259 \times 10^{-5})</td>
<td>1</td>
</tr>
<tr>
<td>Slope:Coast</td>
<td>(-1.5818 \times 10^{-9})</td>
<td>1</td>
</tr>
<tr>
<td>Slope:Sol</td>
<td>(3.2098 \times 10^{-6})</td>
<td>0.798475</td>
</tr>
<tr>
<td>Slope:Temp</td>
<td>(2.2061 \times 10^{-6})</td>
<td>0.867517</td>
</tr>
<tr>
<td>Temp:Sol</td>
<td>(2.2515x \times 10^{-7})</td>
<td>0.762734</td>
</tr>
<tr>
<td>Water:Coast</td>
<td>(4.6803 \times 10^{-9})</td>
<td>0.996914</td>
</tr>
</tbody>
</table>

Table 6.5: Interaction estimates and P-values.


6.6 Conclusions

Previous attempts by researchers to model the distribution of Pā in New Zealand have typically not addressed or accounted for potential confounders. In this chapter we employ a different approach to modelling Pā in order to reduce the confounding effect of population density that we have identified. We have achieved this by instead modelling the relative risk surface estimated in Chapter 4 (Figure 4.8). After fitting both single and multiple covariate models, distance from the coast is the only covariate that exhibits significance in any models tested, and this significance persists for all but the largest bandwidth used. However, we should note as a caveat to this analysis that it is possible that there is some other environmental factor that we have not accounted for that can explain the changes in the relative risk of Pā.

It is difficult to say why the density of Pā may be greater than that of non-Pā sites further inland as there are many differences between the coast and the interior besides proximity to the ocean. However, it is interesting as Pā at first glance appear to tend strongly towards the coast. This is further evidence of the confounding effect of population on the Pā density.

In Chapter 4 we discussed how the density ratio estimate did not support a direct relationship between Pā and agriculture as there are proportionately few Pā in the area of greatest environmental potential. We suggested an alternative hypothesis where Pā have an inverse relationship with arable lands. In this scenario Pā were relatively abundant through the central North Island because arable land is less plentiful, hence creating a greater impetus to protect it. However, the results of this analysis indicate that the probability of observing a Pā compared to another site type is not higher in any locale where the environment is more favourable or unfavourable in terms of agricultural success (temperature, soil drainage, solar radiation). Similar conclusions can be drawn from the other two resources tested: ridges (hunting), and fresh water bodies (fishing and drinking water): the ratio of Pā sites to non-Pā sites is not positively or negatively affected by the abundance of freshwater or the average slope in any given area. These results suggest that in general Pā are not as resource oriented as researchers have previously suggested.

Although the environmental correlates tested in this chapter can be used to predict the locations of Pā as demonstrated by Leathwick (2000) they cannot, with the exception of distance to coast, be used to model the discrepancies between the distribution of Pā and the distribution of other site types in New Zealand. In other words the environmental correlates predict the probability of finding evidence of
any human activity at all, rather than Pā specifically. This further supports that the perceived relationship between these factors and the distribution of Pā is indeed affected by the confounding effects of population density to a certain extent.

If the areas where Pā are more or less dense than other site types cannot be explained by environmental correlates (aside from distance from coast) then these differences are likely driven by differences in social organisation between these regions, or by differences in the use or function of Pā.
Chapter 7

Discussion and Conclusions

At the outset of this thesis we proposed three research questions that have framed the analyses undertaken as part of this study. In this chapter we attempt to answer these research questions and based on our findings offer directions for future research on this subject.

1. What is the density distribution and clustering pattern of Pā sites in New Zealand?

Through the use of kernel density estimation we have produced a statistically supported picture of the distribution of Pā in the two major islands of Aotearoa. These estimates agree closely with the conclusions drawn by other researchers; Pā are most dense in the north of the North Island with the density decreasing gradually southwards. A similar pattern is seen in the South Island with densities being highest in the north-east corner of the island and decreasing southwards. These match closely the three geographical zones discussed in this text: namely, Iwitiini, Waenganni and Te Wāhi Poumanu (Figure 2.3). However, this distribution pattern also closely matches the distribution of the pre-European Māori population, and therefore we argue that inferences from raw density estimates may be limited as behaviours determining the locations of Pā may be obscured. Instead, by comparing the density of Pā with that of other site types via the relative risk function, we have discovered new patterns in the distribution of Pā in the North Island of New Zealand that have hitherto gone unexplored. While the relative risk pattern of Pā in the South Island matches closely the density, the relative risk for the North Island has revealed a very different pattern; in contrast to the straightforward north-south gradient pattern of Pā density, the relative density of Pā compared to non-Pā is significantly low through the Auckland region, and
significantly high through the central North Island.

As for the clustering pattern, by utilising the inhomogeneous pair-correlation function we have found strong evidence of clusters of Pā that repel each other after accounting for other sources of heterogeneity in the Pā point pattern, such as population and geography. These clusters appear to be somewhere between 7 and 20 km in diameter. Repeating this test for the significant regions identified using the relative risk function, we see similar patterns of clustering through the central and northern regions, with similar cluster sizes and repulsion between these regions. A difference can be seen in the northern region where the PCF exceeds one at larger distances indicating that these small clusters may make up large clusters at greater differences.

2. What are the dominant variables governing the density and clustering of Pā sites across New Zealand?

In this thesis we have chosen to model the relative risk of Pā rather than the distribution of Pā in order to curb the confounding effects of population density. The predictor variables tested were land wetness, mean annual temperature, distance from coast, slope, soil drainage, and mean annual solar radiation. Of the various single and multiple covariate models fitted, the only model that was significant was the single covariate model containing distance from the coast. This model indicates that as the distance from the coast increases, the relative risk of observing a Pā compared to a non-Pā also increases.

As the three variables linked to agriculture, i.e., temperature, solar radiation, and soil drainage, do not explain changes in the relative risk surface, this suggests that people did not build more Pā in places more suited to agriculture than they do in places less suited to agriculture, when compared to other site types. Similar conclusions can be drawn from the other variables tested. Changes in the density of Pā relative to non-Pā cannot be explained by the amount of freshwater in a given area or the slope in that area, indicating that access to freshwater and associated food, and ridges for hunting were not significant factors in determining where Pā were built. Additionally, as the relative risk of Pā has a negative association with the coast then this suggests access to kaimoana was also not a major concern when deciding where to build a Pā.

Our findings suggest that the apparent associations between Pā and agriculture, hunting, and water bodies noticed by other researchers may be an observation bias caused by the fact that more people in general are found in places suited to agriculture. That is, we could not identify significant environmental drivers that
cause people to build more or less Pā relative to their population size. One exception is distance from the coast, where Pā appear to have been built with greater density compared to other sites inland. Therefore, if our results indicate any driver for Pā construction it is the general movement inland. This is an interesting finding as the movement of people inland is a phenomenon that does seem to happen more post 1500 CE, the same time that Pā begin to appear.

3. What are the implications of the identified patterns in terms of social organisation?

As discussed in Chapter 2, inferences about pre-European Māori social organisation based on the distribution of Pā have previously been largely drawn from intra-site or regional studies. Researchers carrying out nation-wide studies of Pā distribution have generally not given social explanations for their findings with perhaps the one major exception being the assertion that Pā are evidence of political tension created by competition for agricultural land to support a growing population. However, the analyses undertaken in this thesis suggest that this relationship has been exaggerated by the confounding effects of the pre-contact Māori population distribution. The patterns we have identified and explored in this thesis are markedly different to those that have previously been discussed by archaeologists. We therefore must consider what these patterns can reveal about the social organisation of Māori from circa 1500 CE to the contact period.

We begin by asking whether any of the theories or conclusions about social organisation used to explain patterns found in small scale or regional studies account for the patterns we have seen in this study. Researchers have suggested that Pā were located in relation to one another and interpreted this as an indication of the hapū level of organisation (Buist 1964, Irwin 1985, Phillips 2000). This model of social organisation posits that most competition and conflict occurred between hapū groups (Prickett 1983). Prickett (1983) and Phillips (2000) both noted a pattern of clusters of small Pā with peripheral large Pā. This was interpreted as representing a social hierarchy and complementary relationship. Prickett suggested this pattern allowed rapid congregation of the population at these large Pā for defence if necessary; such a strategy was adopted during the early 18th century when a northern tauā (war party) was defeated by a combined Ngā Mahanga force at Ngaweka Pā, Taranaki (Prickett 1982, 50).

The evidence of clustering we have seen in our point pattern provides support for this notion. The Pā clusters tend to be fairly small at less than 20 km
in diameter (depending on the bandwidth used - if we allow for more of the heterogeneity to come from other sources then the clusters are smaller) and appear to be fairly distinct groups of Pā as we see strong evidence of repulsion beyond this distance. We interpret this relatively small cluster size as indicating that Pā building was mostly controlled by hapū. The repulsion between clusters suggests a fragmented society where hapū are fairly independent, even within their respective iwi: through the central North Island region and through the North Island in general the repulsion between clusters persists for greater distances. This repulsion may be evidence of tension between hapū. This supports Ballara’s 1979 assertion that the hapū was the main political unit in pre-European Māori society.

Through the northern region, however, we see a slightly different pattern. The inhomogeneous PCF suggests clusters of similar size to the central region as well as repulsion between these clusters, but there is also behaviour at larger distances that suggests a different pattern of social organisation; the grouping of clusters at larger distances may be an indication of the iwi level of organisation as it suggests cohesiveness or oversight at a large scale. This suggests that although hapū are still fairly independent, there is more political centralisation through the northern region than through the central region, or through the North Island in general.

By comparing the Pā of the two relative risk regions we may gain more insight into differences in social organisation that may explain differences in the spatial distribution of Pā between these regions. Research suggests that the major difference between the Pā of these regions is their size. Many archaeologists have noted that Pā throughout Taranaki, Waikato and the Bay of Plenty tend to be smaller. Through the northern region and in Auckland in particular we see more large Pā including huge volcanic cone Pā. This is epitomised by One Tree Hill/Maungakiekie (Figure 7.1). In Walton’s 2006 study of the size of Pā he compiles Pā size information from various settlement pattern analyses of areas across the North Island. Of the groups of Pā measured in the central region, the median area is less than 2,000 m$^2$. In contrast, the groups of Pā measured in the northern region have medians greater than 2,000 m$^2$ in all but one case. Additionally the mean area of Pā in the central region is only 2,964 m$^2$ compared to 4,731 m$^2$ in the northern region.

This size discrepancy between the two regions suggests different levels of social organisation. The large Pā of the north indicate a greater degree of political centralisation. These Pā could hold thousands of people at a time and would have taken many man-hours to build. This would have required a large degree
Figure 7.1: Aerial photograph of One Tree Hill/Maungakiekie Pā. Courtesy of Kevin Jones.
of political organisation, possibly at the iwi level. The smaller Pā of the central north island on the other hand suggest they were built by smaller groups, likely hapū, suggesting a possibly more fragmented society through this region. This is consistent with the patterns and conclusions drawn from our cluster analysis. This hypothesis should be explored further by acquiring more information on Pā size.

Conclusions may also be drawn by looking at how the changes in relative risk correspond to tribal boundaries. It should be noted that iwi boundaries were probably fluid and changed over time so modern maps will only give us a general idea of their locations, but at the scale of our analysis this shouldn’t interfere with our conclusions excessively. It is interesting to note that the area where the relative risk is highest, i.e. Taranaki, is the area where there seems to be the highest density of iwi. This is in contrast to Tāmakimakaurau/Auckland where the relative risk is the lowest in the northern region, and there is only one iwi in a reasonably large area. This suggests different levels of political fragmentation between these areas, which could affect the number of Pā built. In the central region iwi, and therefore hapū, have more neighbours than they do in the northern region, possibly leading to greater political tension and hence a greater impetus to build Pā, be they for warfare, defence against (perceived) threats, displays of mana, or other boundary definition behaviour. This greater tension could explain the repulsion between Pā clusters we see in this region. This may tie in to the finding that locations further inland promote a relatively large amount of Pā construction, and that this might be to do with a need to define cultural boundaries in a way that is different from people living on or near the coast. For a settlement pattern focussed on the coastline perhaps territorial boundaries are linear sections of coast with the coast acting as a natural boundary marker and Pā marking the ends of these linear sections. Inland in contrast, this natural boundary is absent, creating the need for polygonal boundaries of Pā; this is known to be a wider pattern in Eastern Polynesia (e.g. Campbell 2001). These polygonal boundaries therefore would require a greater number of Pā as boundary markers, which could be driving the relatively high density of Pā inland. If these polygonal boundaries are serving similar numbers of people to the coastal boundaries then they may not need to serve as many people, which may be a factor driving the smaller size of Pā inland.

To summarise, the density and clustering patterns of Pā in the North Island of New Zealand suggest that society and by extension Pā were generally centred
around the hapū, which were largely independent of their respective iwi in the central North Island, as opposed to the Northern North Island where the clustering pattern suggests some social organisation at the iwi level. This political centralisation in the north allowed the creation of large Pā compared to the fragmented hapū of the central region who were building relatively small Pā. The differences in the relationship between the density of Pā sites and the density of other sites of human activity may be explained by this size difference: in the northern region a relatively small number of large Pā were being built, and in the central region a large number of small Pā were being built. The respective number of independent social groups, be they iwi or hapū, may have also affected the relative number of Pā if a greater number of neighbours lead to more political tension resulting in a greater need for defence, to display mana, or define boundaries. It is tempting to speculate about the ultimate causes of these differences in social organisation: whether it is simply a function of population size (i.e., some threshold where groups tend to aggregate was crossed in the north) or some unmeasured environmental effect (e.g., edges of favourable zones as noted in Chapter 2). This will require further research, but whatever the case the distribution of Pā appears to be only remotely connected to such drivers. Ultimately what our results reveal is that the distribution of Pā is driven more by social factors than environmental ones.

7.1 Looking to the Future

It should be emphasised that this thesis has looked at general patterns across the whole North Island of New Zealand. At the scale of our analysis small-scale local patterns are likely obscured by the larger overall pattern. We acknowledge that if these analyses were repeated at a smaller scale throughout New Zealand, some areas could show markedly different patterns. Accordingly, the methodologies set out in this text can and should be applied to smaller samples of Pā in order to explore distribution patterns further, and to explore in greater depth the theories of pre-European Māori social organisation discussed in this chapter at local and regional levels. Case studies carried out in the northern and central regions would be of particular interest. In such case studies the size of Pā could be measured and hence taken into account more thoroughly than possible here, allowing for more sophisticated analyses that may reveal more about the interlink between Pā size and social organisation/political centralisation.
Additionally, as more Pā are excavated this will allow further interpretation of the results of this thesis as we gain more information on individual Pā, especially with regards to their function/use. Dating of Pā could also allow for more sophisticated analyses. The currently existing dataset of dates for Pā provides poor coverage of New Zealand, especially in the north where Pā are dense and there are proportionately few dates available. A more representative sample of dates would allow spatio-temporal analysis, which may provide more insight into regional differences and social organisation by identifying contemporary Pā and Pā sequences.

This thesis has presented one of the most sophisticated statistical analyses of Pā site locations to date. With the data available at this time, we have been able to provide some novel insights into the distribution of Pā throughout Aotearoa and delved further into this distribution than previous statistical analyses by accounting for confounders and by exploring the implications of our results in terms of pre-European Māori social organisation.
References Cited


Best, E. (2001), *Notes on the art of war: as conducted by the Maori of New Zealand, with accounts of various customs, rites, superstitions, and pertaining to war, as practiced and believed in by the ancient Maori*, Reed, in association with the Polynesian Society, Auckland.


Diggle, P. J. (2003), Statistical Analysis of Spatial Point Pat, Hodder Arnold, Great Britain.


Firth, R. (1927), ‘Maori hill forts’, Antiquity 1, 66–78.


Grey, G. (1885), *Polynesian mythology and ancient traditional history of the New Zealand race, as furnished by their priests and chiefs*, Brett, Auckland.

REFERENCES CITED

Groube, L. M. (1965), *Settlement patterns in New Zealand prehistory*, Anthropology Department, University of Otago, Dunedin.


Kennedy, J. I. (1969), *Settlement in the South East Bay of Islands, 1772: a study in text-aided field archaeology*, Anthropology Department, University of Otago, Dunedin.


REFERENCES CITED


**URL:** [https://www.R-project.org/](https://www.R-project.org/)


